1996

Structural equation and factor analyses for several populations and longitudinal data

Savas Papadopoulos
Iowa State University

Structural equation and factor analyses for several populations and longitudinal data

by

Savas Papadopoulos,

A Dissertation Submitted to the
Graduate Faculty in Partial Fulfillment of the
Requirements for the Degree of
DOCTOR OF PHILOSOPHY

Department: Statistics
Major: Statistics

Approved:

Signature was redacted for privacy.

In Charge of Major Work

Signature was redacted for privacy.

For the Major Department

Signature was redacted for privacy.

For the Graduate College

Iowa State University
Ames, Iowa
1996

UMI Number: 9635342

TABLE OF CONTENTS

ACKNOWLEDGMENTS vi

1. GENERAL INTRODUCTION 1

1.1 Problem and Literature Review 1

1.2 Dissertation Organization 3

1.3 References 5

2. ON LINEAR LATENT VARIABLE ANALYSIS OF MULTIPLE POPULATIONS 8

2.1 Abstract 8

2.2 Introduction 9

2.3 Model, Parameterization, and Procedure 14

2.4 Theoretical Results 22

2.5 Simulation Study 40

2.6 References 50

3. ASYMPTOTIC ROBUSTNESS FOR THE STRUCTURAL EQUATION ANALYSIS OF SEVERAL POPULATIONS 53

3.1 Abstract 53

3.2 Introduction 54

3.3 Notation and Model 56

3.4 Asymptotic Robustness 61

3.5 Monte Carlo Study 65

3.6 References 68

4. FACTOR ANALYSIS OF LONGITUDINAL DATA 71

4.1 Abstract 71

4.2 Introduction 72

4.3 Pseudo-Independence Method 75

4.4 Non-normal Data 83

4.5 Efficiency Comparison 91

4.6 References 96

4.7 Appendix 98

5. CONCLUSION 104

LIST OF TABLES

Table 2.1: Simulation Result for Case 1 (fixed $\xi_j^{(i)}$ $i = 1, 2$) 46

Table 2.2: Simulation Result for Case 2 (correlated $\xi_j^{(i)}$ $i = 1, 2$) 47

Table 2.3: Percentage Relative Bias 48

Table 2.4: Percentage Relative Difference Between Monte Carlo and Asymptotic Standard Errors 48

Table 3.1: Monte Carlo (MC) and Asymptotic Standard Errors 67

Table 4.1: Ratio for the limiting standard deviation (l.s.d.) of $\hat{\beta}_m$ for the first approach 89

Table 4.2: Ratios for the limiting standard deviation (l.s.d.) of $\hat{\beta}_m$ for the second approach 91

Table 4.3: Efficiency of the PI method in comparison to the full likelihood method 94

LIST OF FIGURES

Figure 2.1: Histograms and Non-parametric Density Estimates 49

Figure 4.1: Efficiency of the PI Method in Comparison to the Full-Likelihood Method 95

ACKNOWLEDGMENTS

I would like to thank my major professor Dr. Yasuo Amemiya for his help, suggestions, and uncountable corrections that led my dissertation in its final form. My collaboration with Dr. Yasuo Amemiya was very important and useful for my statistical knowledge and experience that I gained the last two years.

I am indebted to the Survey Section of the Statistical Laboratory for which I was working as a research assistant for four years, to the Department of Statistics for providing me teaching assistantship the last semesters of my graduate studies, and to the Educational Testing Service that gave me the opportunity to participate in the ETS's 1994 Summer Program in Research for Graduate Students.

Definitely, I would like to thank all my close relatives for their support and love. I sincerely want to thank Paraskeui Papadopoulou (mother), Ifikratis Papadopoulos (father), Kiki Papadopoulou (sister), George Korpetis (uncle) and his family, Theodoros Papadopoulos (uncle) and his family, and Panagiota Korpeti (grandmother).

I want to express my thanks to Dr. Wayne A. Fuller who was my advisor for three years, Dr. Ming M. Wang who was my mentor during my internship with the Educational Testing Service, and Dr. Dean Isaacson.

I wish to thank the members of my committee: Professors Jay Breidt, Di Cook, David A. Harville, Frederick O. Lorenz, and Stephen Graham Sapp. I also thank professors Takis Papaioannou, Stavros Kourouklis, George Karakostas, and Nikos Marmaridis for their help.

1. GENERAL INTRODUCTION

1.1 Problem and Literature Review

Latent variable modeling has been used widely in behavioral, biological, and social sciences, as well as in economics. Factor analysis was the first latent variable modeling, and it was introduced by Spearman (1904). In factor analysis, several measurements are taken from each individual, and the relationships among the observed variables are explained in terms of a few underlying factors. For example, if the observed variables are scores from tests, the factors may represent abilities. Another popular latent variable method is the LISREL modeling (LInear Structural RELationships) developed by Keesling (1972), Jõreskog (1973, 1977), and Wiley (1973). The LISREL model assesses relationships among the latent variables in addition to the measurement structure. Latent variables are hypothetical variables, representing underlying concepts, such as ability, social class, anxiety or depression level. Since latent variables cannot be observed directly, we observe variables which are believed to be related to the latent variables. The relationships between the observed variables and the latent variables and the relationships among the latent variables constitute the LISREL model. For instance, we may be interested in the relationship between the cognitive ability and the socio-economic status of the pupils. Both the cognitive ability and the socio-economic status are latent variables. The cognitive ability can be assessed by some tests, while the socio-economic status can be evaluated by observed variables such as parent's education and occupation, and family income. For a general introduction to factor analysis and the LISREL modeling, see, Bollen (1989), Reyment and Jöreskog (1993), and Basilevsky (1994).

In latent variable analyses, the parameter estimation and the model evaluation procedures require numerical computation. The development of computer packages in the last decades has made the latent variable modeling practical and useful. Popular structural equation modeling packages are LISREL (Jöreskog and Sörbom, 1989), EQS (Bentler, 1989), PROC CALIS and PROC FACTOR (SAS, Release 6.10), COSAN (Fraser, 1988), and Mx (Neale, 1994).

Most of the statistical packages assume that all variables are normally distributed. In many cases, the factors and the errors are non-normally distributed, and the use of the packages may lead to incorrect results and conclusions. To solve this problem, asymptotic distribution free (ADF) methods were introduced by, e.g., Bentler (1983), Browne (1984), and Muthén (1989). The ADF methods turned out to have computational and statistical problems for most practical sample sizes, since the sample third-order and fourth-order moments are quite variable. In the last fifteen years, the asymptotic robustness of normality-based methods was studied by several researchers. It was found that these methods can be applied to non-normal data. For references, see Amemiya (1985, 1986), Amemiya, Fuller and Pantula (1987), Browne (1987), Shapiro (1987), Browne and Shapiro (1988), Anderson (1987, 1989), Anderson and Amemiya (1988), Amemiya and Anderson (1990), Browne (1990), Mooijaard and Bentler (1991), Satorra and Bentler (1990), and Satorra (1992, 1993a, 1993b).

The asymptotic robustness of the normality-based methods for multiple populations was examined by Satorra (1993a, 1993b). He considered a model with non-normal random latent variables assuming independent populations and finite fourth-order moments for the non-normal variables. In this dissertation, it is shown that the normality-based methods are robust for latent variable models with fixed and non-normal variables assuming only finite second-order moments for the non-normal variables and allowing some kind of dependency over populations.

Latent variables models can be be fitted to balanced longitudinal data very easily. See Jöreskog and Sörbom (1989). In practice, data are often unbalanced. Then, it may be difficult to estimate all the parameters based on the full-likelihood, especially when the degree of unbalancedness is large. In this document, an efficient method is proposed for unbalanced longitudinal data. The method can be implemented easily using the existing statistical packages, and is particularly useful for non-normal samples.

1.2 Dissertation Organization

This dissertation consists of three papers. The topics and results in the three papers are summarized here.

The first paper discusses linear latent variable analysis of multiple populations. The mean and covariance structures of a general latent variable model for multiple populations are considered. The unknown parameters are estimated by the maximum normal likelihood estimation method. The limiting distribution of the parameter estimators is derived under general assumptions. It is shown that the limiting distribution of the estimators for most important parameters is common under a wide range of assumptions. It is also shown that this result holds for correlated populations. A simulation study is reported using an econometric model with two correlated populations and with fixed and non-normal variables.

In the second paper, augmented-moment structures are considered under four different sets of distributional assumptions for the latent variables. A part of the limiting covariance matrix of the parameter estimator is shown to be identical under the four sets of assumptions. One use of this result is the computation of the correct standard errors by obtaining them under an incorrect but simpler set of assumptions. A simulation study is conducted for an errors-in-variables regression model in two populations under the four sets of assumptions.

The third paper proposes a new statistical analysis method for unbalanced longitudinal data, when the use of the full likelihood or time series approaches are difficult. The proposed method uses a reduced form of the likelihood, and has some good statistical properties. The new method is simple, can be executed by the existing computer packages, and is useful for cases involving non-normal factors and errors. The new method's efficiency relative to the full likelihood method for balanced data, where the latter can be implemented, is numerically shown to be very high.

1.3 References

Amemiya, Y (1985). On the goodness-of-fit tests for linear statistical relationships. *Technical Report No. 10, Econometric Workshop, Stanford University* Stanford, California.

Amemiya, Y (1986). Multivariate functional and structural relationships with general error covariance structure. *Preprint Series No. 86-3, Department of Statistics, Iowa State University,* Ames, Iowa.

Amemiya, Y., Fuller W.A. and Pantula, S.G. (1987). The asymptotic distributions of some estimators for a factor analysis model. *Journal of Multivariate Analysis*, **22**, 51-64.

Amemiya, Y. and Anderson, T.W. (1990). Asymptotic chi-square tests for a large class of factor analysis models. *The Annals of Statistics*, **18**, 1453-1463.

Anderson, T.W. (1987). Multivariate linear relations. *Proceedings of the Second International Conference in Statistics,* edited by Pukkila T. and Puntanen S., University of Tampere, Finland. 9-36.

Anderson, T.W. (1989). Linear latent variable models and covariance structures. *Journal of Econometrics*, **41**, 91-119.

Anderson, T.W. and Amemiya, Y.(1988). The asymptotic normal distribution of estimators in factor analysis under general conditions. *The Annals of Statistics*, **16**, 759-771.

Basilevsky, A. (1994). *Statistical factor analysis and related methods.* John Wiley and Sons, New York.

Bentler, P.M. (1983). Some contributions to efficient statistics in structural models: Specification and estimation of moment structures. Psychometrika, **48**, 493-517.

Bentler, P.M. (1989). *EQS Structural Equations Program Manual.* BMDP Statistical Software, Inc., Los Angeles.

Browne, M.W. (1984). Asymptotically distribution-free methods for the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology*, **37**, 62-83.

Browne, M.W. (1987). Robustness in statistical inference in factor analysis and related models. *Biometrika*, **74**, 375-384.

Browne, M.W. (1990). Asymptotic robustness of normal theory methods for the analysis of latent curves. *Contemporary Mathematics*, **112**, 211-225.

Browne, M.W. and Shapiro, A. (1988). Robustness of normal theory methods in the analysis of linear latent variate models. *British Journal of Mathematical and Statistical Society*, **41**, 193-208.

Bollen, P.A. (1989). *Structural Equations with Latent Variables.* John Wiley and Sons, New York.

Fraser, C. (1988). *COSAN User's Guide.* Unpublished documentation, Center for Behavioral Studies in Education, University of New England, Armidale NSW Australia 2351.

Jöreskog, K. (1973). A general method for estimating a linear structural equation system. In A. S. Goldberger and O. D. Duncan, eds., *Structural Equation Models in the Social Sciences.* New York: Academic Press, pp 85-112.

Jöreskog, K. (1977). Structural equation models in the social sciences: specification, estimation and testing. In P. R. Krishnaiah, ed., *Applications of Statistics.* Amsterdam: North-Holland, pp. 265-287.

Jöreskog, K. and Sörbom, D. (1989). *LISREL 7; A Guide to the Program and Applications.* 2nd ed., SPSS INC., Chicago.

Mooijaart, A. and Bentler, P.M. (1991). Robustness of normal theory statistics in structural equation models. *Statistica Neerlandica* **45**, 159-171.

Muthén, B. (1989). Multiple group structural modeling with non-normal continuous variables. *British Journal of Mathematical and Statistical Psychology*, **42**, 55-62.

Keesling, J. W. (1972). Maximum Likelihood Approaches to Causal Analysis. Ph.D. dissertation. Department of Education: University of Chicago.

Neale, M.C. (1994). *Mx: Structural modeling.* Box 710 MCV, Richmond, VA 23298: Department of Psychiatry. 2nd Edition.

Reyment, R. and Jöreskog (1993). *Applied Factor Analysis in the Natural Sciences.* Cambridge, University Press.

SAS Institute Inc. (1990), *SAS/STAT User's Guide, Version 6, Fourth Edition, Volume 1,* Cary, NC: SAS Institute Inc.

Satorra, A. (1992). Asymptotic robust inferences in the analysis of mean and covariance structures. *Sociological Methodology,* edited by Marsden, P.V., 249-278.

Satorra, A. (1993a). Multi-sample analysis of moment-structures: Asymptotic validity of inferences based on second-order moments. *Statistical Modeling and Latent Variables,* edited by Haagen, K., Bartholomew, D.J. and Deistler, M.. Amsterdam: Elsevier. Forthcoming.

Satorra, A. (1993b). Asymptotic robust inferences multi-sample analysis of augmented-moment structures. *Multivariate Analysis: Future Directions 2,* edited by Cuadras, C.M. and Rao, C.R., 211-229.

Satorra, A and Bentler, P.M. (1990). Model conditions for asymptotic robustness in the analysis of linear relations. *Computational Statistics and Data Analysis,* **10**, 235-249.

Shapiro, A. (1987). Robustness properties of the MDF analysis of moment structures. *South African Statistical Journal,* **21**, 39-62.

Spearman, C. (1904). General intelligence, objectively determined and measured. *American Sociological Review,* **15**, 201-293.

Wiley, D. E. (1973). The identification problem for structural equation models with unmeasured variables. In A. S. Goldberger and O. D. Duncan, eds., *Structural Equation Models in the Social Sciences.* New York: Academic Press, 69-83.

2. ON LINEAR LATENT VARIABLE ANALYSIS OF MULTIPLE POPULATIONS

A paper to be submitted to the Annals of Statistics [1]

Savas Papadopoulos and Yasuo Amemiya

2.1 Abstract

Latent variable or structural equation modeling is used heavily in applications, especially in social and behavioral sciences. Since the normality based model fitting procedures are simple and widely available, and since such procedures are often applied to non-normal or non-random sample data, it is important to investigate the appropriateness of such practice and to suggest simple remedies. This paper addresses these issues for the analysis of multiple populations. For a very general class of latent variable models, a particular parameterization is proposed for meaningful and interpretable analysis of several populations. It is shown that under this parameterization the large-sample statistical inferences based on the assumption of normal and independent populations are valid for virtually any non-normal and dependent

[1] *Key words:* Mean and covariance structure, structural equation modeling, fixed and non-normal factors, asymptotic robustness.

populations. This result is also shown to be valid when some latent variables are treated as fixed instead of random, or when multi-populations in fact correspond to a group of individuals measured over several time points longitudinally. Simulation studies are conducted to verify the theoretical results and assess the use of asymptotic results in finite samples.

2.2 Introduction

Latent variable analysis has been widely used in social, behavioral, and economic sciences, and its applications to medical and business areas are becoming increasingly popular. Structural equation (LISREL) analysis, factor analysis, and errors-in-variables regression are examples of latent variable analysis. In latent variable models, underlying subject-matter concepts are represented by unobservable latent variables, and their relationships with each other and with the observed variables are specified. The models that express observed variables as a linear function of latent variables are extensively used, because of their simple interpretation and the existence of computer packages such as EQS (Bentler (1989)), LISREL (Jöreskog and Sörbom (1989)), and SAS (PROC's CALIS and FACTOR (1990)). The existing computer packages assume that all the variables are normally distributed. The normality and linearity assumptions make the analysis and the interpretation simple, but their applicability in practice is often questionable. In fact, it is rather common in many applications to use the normality-based standard errors and model-fit test procedures when observed variables are highly discrete, bounded, skewed, or obviously non-normal. Thus, it is of practical and theoretical interest to examine the extent of the validity of the normality-based inference procedures for non-normal data, and to explore possible ways to parameterize and formulate a model to attain the wide applicability. In the structural equation analysis literature, this type of research is often referred to as asymptotic robustness study. Most existing results on this topic have been for a single sample from one population. This paper addresses the problem for multiple samples or multiple populations, and provides a unified and comprehensive treatment of the so-called asymptotic robustness. The emphasis is not in extending the single population case results, but in suggesting proper parameterization and modeling leading to practical usefulness in terms of the asymptotic robustness as well as meaningful interpretation.

A general linear latent variable model for a multivariate observation vector $\mathbf{z}$, similar to those considered by Anderson (1987, 1989), Browne and Shapiro (1988), and Satorra (1992), is

$$\mathbf{z} = \mathbf{b}(\boldsymbol{\tau}) + \mathbf{B}_1(\boldsymbol{\tau})\mathbf{f}_1 + \mathbf{B}_2(\boldsymbol{\tau})\mathbf{f}_2 + \cdots + \mathbf{B}_L(\boldsymbol{\tau})\mathbf{f}_L, \tag{2.1}$$

where $\mathbf{f}_\ell$, $\ell = 1, 2, \cdots, L$ are independent latent vectors with means $\mathbf{0}$ and unrestricted covariance matrices $\boldsymbol{\Phi}_\ell$, and $\mathbf{b}(\boldsymbol{\tau})$ and $\mathbf{B}_\ell(\boldsymbol{\tau})$, $\ell = 1, 2, \cdots, L$, are functions of an unknown parameter $\boldsymbol{\tau}$. A common approach to verifying the identification and fitting the model is to assume hypothetically that all $\mathbf{f}_\ell$'s are normally distributed and to concentrate on the first two moments of the observed vector $\mathbf{z}$. The issue for the so-called asymptotic robustness study is to assess the validity of such procedures based on the assumed normality, in terms of inference for unknown parameters and testing the model fit, for a wide class of distributional assumptions on $\mathbf{f}_\ell$'s. It turns out that the type of parameterization used for model (2.1), restricting the coefficient $\mathbf{B}_\ell(\boldsymbol{\tau})$ but keeping the variances $\boldsymbol{\Phi}_\ell$ of the non-normal latent variables $\mathbf{f}_\ell$ unrestricted, plays a key role in the study.

One of the most popular latent variable models used in social and behavioral sciences is the structural equation (LISREL) model, which also includes, as special cases, first and second order factor analysis, measurement error models, and simultaneous equations with errors-in-variables. The LISREL model with mean structures (See, e.g., Jöreskog and Sörbom(1989) or Bollen (1989)) consists of

$$\begin{aligned}
\boldsymbol{\eta} &= \boldsymbol{\alpha} + \mathbf{B}\boldsymbol{\eta} + \boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\zeta}, \\
\mathbf{y} &= \boldsymbol{\gamma}_{\mathbf{y}} + \boldsymbol{\Lambda}_{\mathbf{y}}\boldsymbol{\eta} + \boldsymbol{\epsilon}, \\
\mathbf{x} &= \boldsymbol{\gamma}_{\mathbf{x}} + \boldsymbol{\Lambda}_{\mathbf{x}}\boldsymbol{\xi} + \boldsymbol{\delta},
\end{aligned}$$

where the $d_{\mathbf{y}} \times 1$ $\mathbf{y}$ and $d_{\mathbf{x}} \times 1$ $\mathbf{x}$ are observable, $\boldsymbol{\eta}$ and $\boldsymbol{\xi}$ are underlying latent factors, $\boldsymbol{\zeta}$, $\boldsymbol{\epsilon}$, and $\boldsymbol{\delta}$ are unobservable error vectors, and the last two equations specify measurement models. Assume that $E\{\boldsymbol{\xi}\} = \mathbf{k}$, $E\{\boldsymbol{\zeta}\} = \mathbf{0}$, $E\{\boldsymbol{\epsilon}\} = \mathbf{0}$, $E\{\boldsymbol{\delta}\} = \mathbf{0}$, and $\boldsymbol{\xi}$, $\boldsymbol{\zeta}$, ϵ_m, $m = 1, 2, \cdots, d_{\mathbf{y}}$, and δ_m $m = 1, 2, \cdots, d_{\mathbf{x}}$, are independent non-normal variables with unrestricted variances, where ϵ_m and δ_m are the m^{th} components of $\boldsymbol{\epsilon}$ and $\boldsymbol{\delta}$, respectively. Then, the LISREL model, can be written as a special case of (2.1) by writing

$$\mathbf{z} = \begin{pmatrix} \mathbf{y} \\ \mathbf{x} \end{pmatrix} = \begin{pmatrix} \mathbf{b}_{\mathbf{y}} \\ \mathbf{b}_{\mathbf{x}} \end{pmatrix} + \begin{pmatrix} \boldsymbol{\Lambda}_{\mathbf{y}}(I-\mathbf{B})^{-1}\boldsymbol{\Gamma} \\ \boldsymbol{\Lambda}_{\mathbf{x}} \end{pmatrix} (\boldsymbol{\xi} - \mathbf{k}) + \begin{pmatrix} \boldsymbol{\Lambda}_{\mathbf{y}}(I-\mathbf{B})^{-1} \\ \mathbf{0}_{d_{\mathbf{x}} \times d_{\mathbf{x}}} \end{pmatrix} \boldsymbol{\zeta}$$

$$+ \begin{pmatrix} \mathbf{I}^{(\mathbf{y})}_{\bullet 1} \\ \mathbf{0}_{d_{\mathbf{x}}} \end{pmatrix} \epsilon_1 + \cdots + \begin{pmatrix} \mathbf{I}^{(\mathbf{y})}_{\bullet d_{\mathbf{y}}} \\ \mathbf{0}_{d_{\mathbf{x}}} \end{pmatrix} \epsilon_{d_{\mathbf{y}}} + \begin{pmatrix} \mathbf{0}_{d_{\mathbf{y}}} \\ \mathbf{I}^{(\mathbf{x})}_{\bullet 1} \end{pmatrix} \delta_1 + \cdots + \begin{pmatrix} \mathbf{0}_{d_{\mathbf{y}}} \\ \mathbf{I}^{(\mathbf{x})}_{\bullet d_{\mathbf{x}}} \end{pmatrix} \delta_{d_{\mathbf{x}}},$$

where $\mathbf{b}_{\mathbf{y}} = \boldsymbol{\gamma}_{\mathbf{y}} + \boldsymbol{\Lambda}_{\mathbf{y}}(I-\mathbf{B})^{-1}\boldsymbol{\alpha} - \boldsymbol{\Lambda}_{\mathbf{y}}(I-\mathbf{B})^{-1}\boldsymbol{\Gamma}\mathbf{k}$, $\mathbf{b}_{\mathbf{x}} = \boldsymbol{\gamma}_{\mathbf{x}} - \boldsymbol{\Lambda}_{\mathbf{x}}\mathbf{k}$, and $\mathbf{I}^{(\mathbf{y})}_{\bullet m}$ is the m^{th} column of the $d_{\mathbf{y}} \times d_{\mathbf{y}}$ identity matrix $\mathbf{I}$. This is in the form of model (2.1) with

$L = 2 + d_{\mathbf{y}} + d_{\mathbf{x}}$ and $\boldsymbol{\tau}$ containing unknown parameters in $\boldsymbol{\gamma}_{\mathbf{y}}$, $\boldsymbol{\gamma}_{\mathbf{x}}$, $\boldsymbol{\Lambda}_{\mathbf{y}}$, $\boldsymbol{\Lambda}_{\mathbf{x}}$, $\boldsymbol{\alpha}$, $\mathbf{B}$, $\boldsymbol{\Gamma}$, and $\mathbf{k}$.

In this paper, the multiple population cases of model (2.1) are discussed. Latent variable analysis of multiple populations was discussed by Jöreskog (1971), Lee and Tsui (1982), Muthén (1989), and Satorra (1993a, 1993b). Consider multivariate samples from several populations, where a version of (2.1) holds for each population. Then, the interest may be in making inferences about the similarities and differences among populations. If similar variables are measured from each population, then the parameters or characteristics associated with a measurement process are assumed to be common over the samples. Even in such a case, some of the latent variables being measured can have different characteristics (in terms of different distributional parameters) over populations. The existing computer packages, LISREL and EQS, can analyze multiple populations simultaneously under the assumption that the populations are independent. Another type of the multi-population problem is concerned with the so-called correlated populations. The multiple samples may in fact come from the same population over different time periods (multivariate repeated measures) or may be spatially correlated. See, e.g., Papadopoulos and Amemiya (1995). Such correlated or dependent population (or sample) cases have not been fully discussed in the literature, and will be treated in this paper. For both types of multi-population problems, this paper considers simple model fitting and testing procedures that can be readily carried out using the existing packages. We discuss a general multi-population model possibly containing fixed, non-normal, and normal components, and introduce a way to formulate and parameterize the model so that the multi-population analysis can be conducted and interpreted meaningfully in practice, and so that the so-called asymptotic robustness is achieved for inferences concerning parameters and model fit.

The so-called asymptotic robustness of normal-based methods for the latent variable analysis has been extensively studied in the last ten years. For, exploratory (unrestricted) factor analysis, Amemiya, Fuller, and Pantula (1987) proved that the limiting distribution of some estimators is the same for fixed, non-normal, and normal factors under the assumption that the errors are normally distributed. Amemiya (1986) treated functional and structural relationships with error covariance matrix as a function of an unknown parameter vector. The robustness of goodness-of-fit tests was studied by Amemiya (1985). Browne (1987) showed that the above results hold for a more general class of latent variable models assuming finite eighth moments for the factors and normal errors. Anderson and Amemiya (1988), and Amemiya and Anderson (1990) extended the above results to confirmatory factor analysis and non-normal errors; they assume finite second moments for the factors and errors. Browne and Shapiro (1988) introduced a general linear model, and followed a different approach from Amemiya and Anderson which forces the observed variables to have finite fourth moments. Considering the model of Browne and Shapiro, Anderson (1987, 1989) included non-stochastic latent variables, and assumed only finite second moments for the non-normal latent variables. Latent variable models with mean and covariance structures were studied by Browne (1990) and Satorra (1992). Satorra (1993a, 1993b, 1994) considered the multi-sample analysis of augmented-moment structures. Additional studies on the asymptotic robustness of latent variable analysis were conducted by Shapiro (1987), Mooijaard and Bentler (1991), and Satorra and Bentler (1990).

Asymptotic distribution-free (ADF) methods for the latent variable analysis were proposed to deal with non-normal data. See, e.g., Bentler (1983), Browne (1984), and Muthén (1989). The ADF methods turned out to be problematic in practice, since the fourth-order sample moments are very variable. See, e.g., Chou, Bentler, and Satorra (1991) and Muthén and Kaplan (1992).

In this paper, mean and covariance structures are considered for a general multi-population model which contains fixed, normal and non-normal variables; some of the non-normal variables are allowed to be correlated over populations. We use the approach of Anderson and Amemiya (1988) to show that the normal-based methods are applicable for non-normal and non-random data assuming finite second-order moments.

Section 2.3 introduces a general multi-population model and associated assumptions and parameterizations. Model fitting and checking procedures are also defined in Section 2.3. The theoretical results are derived and discussed in Section 2.4. Section 2.5 reports results from simulation studies.

2.3 Model, Parameterization, and Procedure

Consider I populations. Suppose that $n^{(i)}$ individuals are sampled from the i-th population, $i = 1, 2, \cdots, I$, and that $p^{(i)}$ measurements are taken from each sampled individual in the i-th population. Denote the multi-sample data set,

$$\{\mathbf{z}_j^{(i)} :\ i = 1, 2, \cdots, I;\ j = 1, 2, \cdots, n^{(i)}\}, \tag{2.2}$$

where $\mathbf{z}_j^{(i)}$ is the $p^{(i)} \times 1$ measurement vector from the j-th individual in the i-th population. We consider a very general latent variable model that includes models widely used in single population cases and covers a large class of distributional situations in one form. To cover various distributional settings, it is convenient to assume that the observed vector $\mathbf{z}_j^{(i)}$ can be written as a linear combination of $L^{(i)}+2$ independent latent vectors, $i=1,2,\cdots,I$, and that the latent vectors can be divided into three groups; a random vector $\mathbf{f}_{0j}^{(i)}$ assumed to be normally distributed, $L^{(i)}$ random vectors $\mathbf{f}_{\ell j}^{(i)}$, $\ell=1,\cdots,L^{(i)}$, assumed to have unspecified or non-normal distributions, and a fixed vector $\mathbf{f}_{L^{(i)}+1,j}^{(i)}$. The dimension of $\mathbf{f}_{\ell j}^{(i)}$ is $q_\ell^{(i)}\times 1$, $\ell=0,1,\cdots,L^{(i)}+1$. The model can be written as

$$\mathbf{z}_j^{(i)}=\mathbf{b}^{(i)}+\mathbf{B}_0^{(i)}\mathbf{f}_{0j}^{(i)}+\mathbf{B}_1^{(i)}\mathbf{f}_{1j}^{(i)}+\cdots+\mathbf{B}_{L^{(i)}}^{(i)}\mathbf{f}_{L^{(i)}j}^{(i)}+\mathbf{B}_{L^{(i)}+1}^{(i)}\mathbf{f}_{L^{(i)}+1,j}^{(i)}, \qquad (2.3)$$

for $i=1,2,\cdots,I$, $j=1,2,\cdots,n^{(i)}$, where the $p^{(i)}\times 1$ $\mathbf{b}^{(i)}$ and $p^{(i)}\times q_\ell^{(i)}$ $\mathbf{B}_\ell^{(i)}$, $\ell=0,1,\cdots,L^{(i)}+1$, contain unknown parameters. Note that the sample size $n^{(i)}$, the number of measured variables $p^{(i)}$, and the number of latent vectors $L^{(i)}$ generally differ over populations (depend on i). This generality of the model allows us to deal with cases where slightly different variables are measured from different populations with possibly different structures. The associated assumptions are summarized in a set as follows:

Assumption 1 *i) For a given individual j,* $\{\mathbf{f}_{\ell j}^{(i)}\ i=1,2,\cdots,I;\ \ell=0,1,\cdots,L^{(i)}\}$ *are mutually independent with* $E\{\mathbf{f}_{\ell j}^{(i)}\}=\mathbf{0}$ *and* $Var\{\mathbf{f}_{\ell j}^{(i)}\}=\mathbf{\Phi}_\ell^{(i)}$.

ii) For a given population $i=1,2,\cdots,I$ *and for a given latent random vector index* $\ell=0,1,\cdots,L^{(i)}$, $\{\mathbf{f}_{\ell j}^{(i)},\ \ j=1,2,\cdots,n^{(i)}\}$ *are independently and identically distributed.*

iii) $\mathbf{f}_{0j}^{(i)}$ *are normally distributed for all i and j.*

iv) $\{\mathbf{f}^{(i)}_{L^{(i)}+1,j},\ \ i=1,2,\cdots,I;\ j=1,2,\cdots,n^{(i)}\}$ *are fixed constants satisfying*

$$\lim_{n^{(i)}\to\infty} \bar{\mathbf{f}}^{(i)}_{L^{(i)}+1}(n^{(i)}) \ = \ \lim_{n^{(i)}\to\infty} \frac{1}{n^{(i)}}\sum_{j=1}^{n^{(i)}} \mathbf{f}^{(i)}_{L^{(i)}+1,j} = \boldsymbol{\varphi}^{(i)\infty}_{L^{(i)}+1}, \tag{2.4}$$

$$\begin{aligned}\lim_{n^{(i)}\to\infty} \boldsymbol{\Phi}^{(i)}_{L^{(i)}+1}(n^{(i)}) \ &= \ \lim_{n^{(i)}\to\infty} \frac{1}{n^{(i)}-1}\sum_{j=1}^{n^{(i)}} (\mathbf{f}^{(i)}_{L^{(i)}+1,j} - \bar{\mathbf{f}}^{(i)}_{L^{(i)}+1})(\mathbf{f}^{(i)}_{L^{(i)}+1,j} - \bar{\mathbf{f}}^{(i)}_{L^{(i)}+1})' \\ &= \ \boldsymbol{\Phi}^{(i)\infty}_{L^{(i)}+1}, \end{aligned} \tag{2.5}$$

where $\boldsymbol{\Phi}^{(i)\infty}_{L^{(i)}+1}$ *is a positive definite matrix.*

v) For $i = 1,2,\cdots,I$, $\mathbf{b}^{(i)} = \mathbf{b}^{(i)}(\boldsymbol{\tau})$, $\mathbf{B}^{(i)}_\ell = \mathbf{B}^{(i)}_\ell(\boldsymbol{\tau})$ $(\ell = 0,1,2,\cdots,L^{(i)}+1)$, *and the covariance matrix* $\boldsymbol{\Phi}^{(i)}_0 = \boldsymbol{\Phi}^{(i)}_0(\boldsymbol{\tau})$ *of normal* $\mathbf{f}^{(i)}_{0j}$ *are functions of a parameter vector* $\boldsymbol{\tau}$ *of dimension* $d_{\boldsymbol{\tau}} \times 1$ *in an open parameter space* $\boldsymbol{\Gamma}_{\boldsymbol{\tau}}$. *It is assumed that* $\boldsymbol{\Phi}^{(i)}_0(\boldsymbol{\tau})$ *is positive definite for all* $\boldsymbol{\tau}$ *in* $\boldsymbol{\Gamma}_{\boldsymbol{\tau}}$. *For* $i = 1,2,\cdots,I$, *and* $\ell = 1,2,\cdots,L^{(i)}$, $\boldsymbol{\Phi}^{(i)}_\ell$ *are unrestricted positive definite matrices.*

Under Assumption 1, all $\mathbf{f}^{(i)}_{\ell j}$'s are independent. The treatment of cases with dependent populations or samples will be discussed later. All normally distributed latent variables are included in $\mathbf{f}^{(i)}_{0j}$, and their distributions may possibly be related through $\boldsymbol{\tau}$ over populations, $i = 1,2,\cdots,I$. Other unspecified or non-normal random latent variables are divided into independent parts $\ell = 1,2,\cdots,L^{(i)}$ with unrestricted covariance matrices. The fixed $\mathbf{f}^{(i)}_{L^{(i)}+1,j}$ can represent a situation where the interest is in the model fitting and estimation only for a given set of individuals and not for the populations. In addition, the fixed $\mathbf{f}^{(i)}_{L^{(i)}+1,j}$ can be used in an analysis conducted conditionally on a given set of $\mathbf{f}^{(i)}_{L^{(i)}+1,j}$ values. Such a conditional analysis may be appropriate when the individuals $j = 1,2,\cdots,n^{(i)}$ do not form a random sample from the i-th population and/or when a component of $\mathbf{z}^{(i)}_j$ represents some dependency over I populations. For example, the I populations may actually correspond to a single population at I different time points. The use of such a conditional argument will also play an important role in our theoretical development as well. With $\mathbf{f}^{(i)}_{L^{(i)}+1,j}$ being latent and fixed, the limit of the unobservable sample mean $\boldsymbol{\varphi}^{(i)\infty}_{L^{(i)}+1}$ is assumed unknown and unrestricted, and the limit of the unobservable sample covariance matrix $\boldsymbol{\Phi}^{(i)\infty}_{L^{(i)}+1}$ is assumed to be an unknown and unrestricted positive definite matrix. All $\mathbf{b}^{(i)}(\boldsymbol{\tau})$ and $\mathbf{B}^{(i)}_{\ell}(\boldsymbol{\tau})$ are expressed in terms of $\boldsymbol{\tau}$ allowing functional relationships over I populations and known or related elements. Even though $\boldsymbol{\tau}$ also appears in $\boldsymbol{\Phi}^{(i)}_0(\boldsymbol{\tau})$, the elements of $\boldsymbol{\tau}$ are usually divided into two groups; one for $\boldsymbol{\Phi}^{(i)}_0(\boldsymbol{\tau})$ and another for $\mathbf{b}^{(i)}(\boldsymbol{\tau})$ and $\mathbf{B}^{(i)}_{\ell}(\boldsymbol{\tau})$. Assumption 1 v) provides a particular identifiable parameterization for model (2.3). For a single population case with $I = 1$, various equivalent parameterization have been used in practice. Some place restrictions on covariance matrices, e.g., by standardizing latent variables, and leave the coefficients unrestricted. The parameterization that leaves the covariance matrices (and mean vectors) of latent variables unrestricted and that places identification restrictions only on the coefficients and intercepts is referred to as the errors-in-variables parameterization. For the single population case, a parameterization with restricted covariance matrices generally has an equivalent errors-in-variables parameterization, and the two parameterizations with one-to-one correspondence lead to an equivalent interpretation. The single sample asymptotic robustness results have shown that the asymptotic standard errors for the parameters in the errors-in-variables formulation computed under the normality assumption are valid for non-normal data, but that the same does not hold under parameterization with restricted covariance matrices. For the multi-sample model (2.3), we will show that the errors-in-variables type parameterization, given in Assumption 1, provides the asymptotic robustness. However, for the multi-sample case, there are other reasons for considering the parameterization specified in Assumption 1 v). As mentioned earlier, a multi-population study is conducted because the populations are thought to be different, but certain aspects of the structure generating data are believed to be common over populations. Suppose that the same or similar measurements are taken from different populations. For example, a similar set of psychological tests may be given to a number of different groups, e.g., two gender groups, groups with different occupations or educational backgrounds, groups in varying socio-economic or cultural environments, or different time points in the growth of a group. The subject matter or scientific interest exists in making inferences about some general assertion that holds commonly for various populations. Such interest is usually expressed as relationships among latent (and observed) variables that hold regardless of the location and variability of the variables. Then, a relevant analysis is to estimate and test the relationships, and to explore the range of populations for which the relationships hold. The parameterization in Assumption 1 v) with unrestricted $\mathbf{\Phi}_\ell^{(i)}$ and generally structured $\mathbf{B}_\ell^{(i)}(\boldsymbol{\tau})$ corresponds very well with the scientific interest of the study, and allows the interpretation consistent with the practical meaning of the problem. Note that $\mathbf{\Phi}_\ell^{(i)}$ are unrestricted covariance matrices and do not have any relationships over i or ℓ, and that $\mathbf{b}^{(i)}(\boldsymbol{\tau})$ and $\mathbf{B}_\ell^{(i)}(\boldsymbol{\tau})$ can have known elements and elements with relationships over i and ℓ. On the other hand, the covariance matrix $\mathbf{\Phi}_0^{(i)}(\boldsymbol{\tau})$ of the normal latent vector $\mathbf{f}_{0j}^{(i)}$ can have restrictions or equality over populations through $\boldsymbol{\tau}$. This gives the generality of model (2.3) with only one normal latent vector, because a block diagonal $\mathbf{\Phi}_0^{(i)}(\boldsymbol{\tau})$ corresponds to a number of independent sub-vectors in the normal $\mathbf{f}_{0j}^{(i)}$. In addition, the possibility of restrictions on $\mathbf{\Phi}_0^{(i)}(\boldsymbol{\tau})$ over populations can also be important in applications. For example, if the same measurement instruments are applied to different samples, then the variances of pure measurement errors may be assumed to be the same over the samples. However, the normal assumption for pure measurement errors is reasonable in most situations, and such errors can be included in $\mathbf{f}_{0j}^{(i)}$. Assumption 1 v) does not rule out latent variable variances and covariances with restrictions across populations, but does require the latent variables with restricted variances to be normally distributed. This requirement is not very restrictive in most applications, as discussed above, but it is needed to obtain the asymptotic robustness results given in the next section. The general form of $\mathbf{b}^{(i)}(\boldsymbol{\tau})$ and the inclusion of fixed latent vector $\mathbf{f}_{L^{(i)}+1,j}^{(i)}$ allow virtually any structure for the means of the observed $\mathbf{z}_j^{(i)}$. Hence, the errors-in-variables type parameterization in Assumption 1 v) can solve the identification problem, provides a general and a convenient way to represent the subject-matter theory and concepts, and produces asymptotic robustness results presented in the next section.

For the multi-sample data $\{\mathbf{z}_j^{(i)}\}$ in (2.2), let

$$\bar{\mathbf{z}}^{(i)} = \frac{1}{n^{(i)}} \sum_{j=1}^{n^{(i)}} \mathbf{z}_j^{(i)}, \quad \mathbf{S}^{(i)} = \frac{1}{n^{(i)} - 1} \sum_{j=1}^{n^{(i)}} (\mathbf{z}_j^{(i)} - \bar{\mathbf{z}}^{(i)})(\mathbf{z}_j^{(i)} - \bar{\mathbf{z}}^{(i)})' \tag{2.6}$$

be the sample mean vector and sample covariance matrix for the i-th population ($i = 1, 2, \cdots, I$). It is assumed that the sample covariance matrices $\mathbf{S}^{(i)}$ are nonsingular with probability one. For a $p \times p$ symmetric matrix $\mathbf{A}$, let $\mathrm{v}(\mathbf{A})$ denote the $p(p+1)/2 \times 1$ vector that includes the elements of $\mathbf{A}$ on or below the diagonal starting with the first column. Define

$$\mathbf{c}^{(i)} = \begin{pmatrix} \bar{\mathbf{z}}^{(i)} \\ \mathrm{v}(\mathbf{S}^{(i)}) \end{pmatrix}, \qquad \mathbf{c} = \begin{pmatrix} \mathbf{c}^{(1)} \\ \vdots \\ \mathbf{c}^{(I)} \end{pmatrix}. \tag{2.7}$$

We consider model fitting and estimation based only on $\mathbf{c}$, because such procedures can be carried out using the existing computer packages, because Assumption 1 does not specify a particular distributional form of observations beyond the first two moments, and because no particular correspondence or relationship between samples is specified in Assumption 1. Let $\boldsymbol{\theta}$ be a $d_{\boldsymbol{\theta}} \times 1$ vector containing all unknown parameters in $E\{\mathbf{c}\} = \boldsymbol{\gamma}(\boldsymbol{\theta})$ under model (2.3) and Assumption 1, and let

$$\boldsymbol{\theta} = \begin{pmatrix} \boldsymbol{\tau} \\ \boldsymbol{\rho} \\ \boldsymbol{\varphi} \end{pmatrix}, \qquad \boldsymbol{\rho} = \begin{pmatrix} \boldsymbol{\rho}^{(1)} \\ \vdots \\ \boldsymbol{\rho}^{(I)} \end{pmatrix}, \qquad \boldsymbol{\rho}^{(i)} = \begin{pmatrix} \mathrm{v}(\boldsymbol{\Phi}_1^{(i)}) \\ \vdots \\ \mathrm{v}(\boldsymbol{\Phi}_{L^{(i)}}^{(i)}) \end{pmatrix},$$

$$\boldsymbol{\varphi} = \begin{pmatrix} \boldsymbol{\varphi}^{(1)} \\ \vdots \\ \boldsymbol{\varphi}^{(I)} \end{pmatrix}, \qquad \boldsymbol{\varphi}^{(i)} = \begin{pmatrix} \bar{\mathbf{f}}_{L^{(i)}+1}^{(i)} \\ \mathrm{v}[\boldsymbol{\Phi}_{L^{(i)}+1}^{(i)}(n^{(i)})] \end{pmatrix}. \tag{2.8}$$

Note that $\boldsymbol{\varphi}$ consists of the unobservable sample moments of the fixed $\mathbf{f}_{L^{(i)}+1}^{(i)}$, and depends on $n^{(i)}$. The dimensions of the vectors $\boldsymbol{\rho}$ and $\boldsymbol{\varphi}$ are $d_{\boldsymbol{\rho}}$ and $d_{\boldsymbol{\varphi}}$, where

$$d_{\boldsymbol{\rho}} = \sum_{i=1}^{I} d_{\boldsymbol{\rho}}^{(i)}, \qquad d_{\boldsymbol{\rho}}^{(i)} = \sum_{\ell=1}^{L^{(i)}} \frac{q_\ell^{(i)}(q_\ell^{(i)}+1)}{2},$$

$$d_{\boldsymbol{\varphi}} = \sum_{i=1}^{I} [q_{L^{(i)}+1}^{(i)} + \frac{q_{L^{(i)}+1}^{(i)}(q_{L^{(i)}+1}^{(i)}+1)}{2}].$$

It is assumed that $d_{\boldsymbol{\theta}} = d_{\boldsymbol{\tau}} + d_{\boldsymbol{\rho}} + d_{\boldsymbol{\varphi}} \leq d_{\mathbf{c}}$, where $d_{\mathbf{c}}$, the dimension of $\mathbf{c}$, is

$$d_{\mathbf{c}} = \sum_{i=1}^{I} d_{\mathbf{c}}^{(i)}, \qquad d_{\mathbf{c}}^{(i)} = p^{(i)} + \frac{p^{(i)}(p^{(i)}+1)}{2}.$$

Under model (2.3) and Assumption 1,

$$\boldsymbol{\mu}^{(i)}(\boldsymbol{\theta}) = E\{\bar{\mathbf{z}}^{(i)}\} = \mathbf{b}^{(i)}(\boldsymbol{\tau}) + \mathbf{B}^{(i)}_{L^{(i)}+1}(\boldsymbol{\tau})\bar{\mathbf{f}}^{(i)}_{L^{(i)}+1}, \tag{2.9}$$

$$\begin{aligned}\boldsymbol{\Sigma}^{(i)}(\boldsymbol{\theta}) = E\{\mathbf{S}^{(i)}\} &= \mathbf{B}^{(i)}_0(\boldsymbol{\tau})\boldsymbol{\Phi}^{(i)}_0(\boldsymbol{\tau})\mathbf{B}^{(i)\prime}_0(\boldsymbol{\tau}) + \sum_{\ell=1}^{L^{(i)}} \mathbf{B}^{(i)}_\ell(\boldsymbol{\tau})\boldsymbol{\Phi}^{(i)}_\ell \mathbf{B}^{(i)\prime}_\ell(\boldsymbol{\tau}) \\ &\quad + \mathbf{B}^{(i)}_{L^{(i)}+1}(\boldsymbol{\tau})\boldsymbol{\Phi}^{(i)}_{L^{(i)}+1}(n^{(i)})\mathbf{B}^{(i)\prime}_{L^{(i)}+1}(\boldsymbol{\tau}). \end{aligned} \tag{2.10}$$

Denote

$$E\{\mathbf{c}\} = \boldsymbol{\gamma}(\boldsymbol{\theta}) = \begin{pmatrix} \boldsymbol{\gamma}^{(1)}(\boldsymbol{\theta}) \\ \vdots \\ \boldsymbol{\gamma}^{(I)}(\boldsymbol{\theta}) \end{pmatrix}, \qquad \boldsymbol{\gamma}^{(i)}(\boldsymbol{\theta}) = \begin{pmatrix} \boldsymbol{\mu}^{(i)}(\boldsymbol{\theta}) \\ \mathrm{v}[\boldsymbol{\Sigma}^{(i)}(\boldsymbol{\theta})] \end{pmatrix}.$$

Let $\boldsymbol{\Gamma}_{\boldsymbol{\theta}}$ be the parameter space for $\boldsymbol{\theta}$ under Assumption 1. Recall that the parameters $\boldsymbol{\varphi}^{(i)}$ in $\boldsymbol{\theta}$ depend on $n^{(i)}$. In $\boldsymbol{\Gamma}_{\boldsymbol{\theta}}$, the parameter spaces for $\bar{\mathbf{f}}^{(i)}_{L^{(i)}+1}$ and $\boldsymbol{\Phi}^{(i)}_{L^{(i)}+1}(n^{(i)})$ are the set of all $q^{(i)}_{L^{(i)}+1} \times 1$ vectors and the set of all $q^{(i)}_{L^{(i)}+1} \times q^{(i)}_{L^{(i)}+1}$ symmetric positive definite matrices respectively. For the estimation of $\boldsymbol{\theta}$, we consider an estimator $\hat{\boldsymbol{\theta}}$ obtained by minimizing over $\boldsymbol{\Gamma}_{\boldsymbol{\theta}}$

$$Q(\boldsymbol{\gamma}(\boldsymbol{\theta}),\ \mathbf{c}) = \sum_{i=1}^{I} \frac{n^{(i)}}{n} Q^{(i)}, \tag{2.11}$$

where $n = n^{(1)} + \cdots + n^{(I)}$ is the total sample size, and

$$\begin{aligned} Q^{(i)}(\boldsymbol{\gamma}^{(i)}(\boldsymbol{\theta}),\ \mathbf{c}^{(i)}) &= \mathrm{tr}\{\mathbf{S}^{(i)}\boldsymbol{\Sigma}^{(i)-1}(\boldsymbol{\theta})\} - \log|\mathbf{S}^{(i)}\boldsymbol{\Sigma}^{(i)-1}(\boldsymbol{\theta})| - p^{(i)} \\ &\quad + [\bar{\mathbf{z}}^{(i)} - \boldsymbol{\mu}^{(i)}(\boldsymbol{\theta})]'\boldsymbol{\Sigma}^{(i)-1}(\boldsymbol{\theta})[\bar{\mathbf{z}}^{(i)} - \boldsymbol{\mu}^{(i)}(\boldsymbol{\theta})]. \end{aligned}$$

The obtained estimator $\hat{\boldsymbol{\theta}}$ is a slight modification of the normal maximum likelihood estimator (MLE). The exact normal MLE can be obtained if $[(n^{(i)} - 1)/n^{(i)}]\mathbf{S}^{(i)}$ is used in place of $\mathbf{S}^{(i)}$. Asymptotic results are equivalent for the two estimators. We consider $\hat{\boldsymbol{\theta}}$ because it can be computed by the existing computer packages. The form of $Q^{(i)}$ corresponds to the so-called mean and covariance structure analysis. But, the existing covariance structure computer packages without mean structure can be used to carry out the minimization of $Q(\boldsymbol{\gamma}(\boldsymbol{\theta}),\ \mathbf{c})$ using a certain technique. See, e.g., the manuals of LISREL and EQS. In the next section, asymptotic distribution results for $\hat{\boldsymbol{\theta}}$ are derived for the broad class of situations.

2.4 Theoretical Results

To derive large sample results for $\hat{\boldsymbol{\theta}}$ minimizing (2.11) under model (2.3) and Assumption 1, we consider the case where all $n^{(i)}$ increase to infinity at a common rate, and use $n_m = \min\{n^{(1)}, \cdots, n^{(I)}\}$ as the index for taking a limit:

Assumption 2 $\lim_{n_m \to \infty} n^{(i)}/n = r^{(i)} > 0$.

Before introducing the notation for the true value of the parameter $\boldsymbol{\theta}$ in (2.8) for model (2.3), note that the $\boldsymbol{\varphi}$ part of $\boldsymbol{\theta}$, corresponding to the unobservable sample moments of the fixed $\mathbf{f}^{(i)}_{L^{(i)}+1,j}$, depends on $n^{(i)}$. Thus, we denote the true value of $\boldsymbol{\varphi}$ for this particular sample by $\boldsymbol{\varphi}_n$. Under Assumption 1 *iv*), the true value $\boldsymbol{\varphi}_n$ has a limit as $n_m \to \infty$, i.e.,

$$\lim_{n_m \to \infty} \boldsymbol{\varphi}_{\mathbf{n}} = \boldsymbol{\varphi}_{\infty},$$

where $\boldsymbol{\varphi}_{\infty}$ consists of $\boldsymbol{\varphi}^{(i)\infty}_{L^{(i)}+1}$ and $\boldsymbol{\Phi}^{(i)\infty}_{L^{(i)}+1}$, $i = 1, 2, \cdots, I$ given in Assumption 1 *iv*). Let $\boldsymbol{\tau}_0$ and $\boldsymbol{\rho}_0$ be the true values of the $\boldsymbol{\tau}$ and $\boldsymbol{\rho}$ parts of $\boldsymbol{\theta}$ in (2.8) associated with the true distribution generating $\mathbf{z}^{(i)}_j$'s under model (2.3) and Assumption 1. Also, let

$$\boldsymbol{\theta}_{\mathbf{n}} = (\boldsymbol{\tau}_0',\ \boldsymbol{\rho}_0',\ \boldsymbol{\varphi}_{\mathbf{n}}')' \tag{2.12}$$

$$\boldsymbol{\theta}_\infty = (\boldsymbol{\tau}_0',\ \boldsymbol{\rho}_0',\ \boldsymbol{\varphi}_\infty')' = \lim_{n_m \to \infty} \boldsymbol{\theta}_{\mathbf{n}} \tag{2.13}$$

denote the true value of $\boldsymbol{\theta}$ and the limiting true value of $\boldsymbol{\theta}$, respectively. Since model (2.3) is very general, and since the parameters are identified only through $\boldsymbol{\gamma}(\boldsymbol{\theta}) = E\{\mathbf{c}\}$, a standard identification condition is needed to obtain large sample properties:

Assumption 3 *For any $\varepsilon > 0$ there exists a $\delta > 0$ such that if $\| \boldsymbol{\gamma}(\boldsymbol{\theta}) - \boldsymbol{\gamma}(\boldsymbol{\theta}_\infty) \| < \delta$ then $\| \boldsymbol{\theta} - \boldsymbol{\theta}_\infty \| < \varepsilon$, where $\| \mathbf{x} \| = \sqrt{\mathbf{x}'\mathbf{x}}$ for a vector $\mathbf{x}$.*

The first theorem gives the consistency of the estimator $\hat{\boldsymbol{\theta}}$ minimizing (2.11) for model (2.3) in the sense that $\mathrm{plim}_{n_m \to \infty} \hat{\boldsymbol{\theta}} = \boldsymbol{\theta}_\infty$.

Theorem 1 *Assume that model (2.3) holds. If Assumptions 1, 2, and 3 hold, then as $n_m \to \infty$*

$$\hat{\boldsymbol{\theta}} \xrightarrow{p} \boldsymbol{\theta}_\infty.$$

Proof: Under model (2.3), $\bar{\mathbf{z}}^{(i)}$ and $\mathbf{S}^{(i)}$, defined in (2.6), satisfy

$$\bar{\mathbf{z}}^{(i)} = \mathbf{b}^{(i)}(\boldsymbol{\tau}_0) + \sum_{\ell=0}^{L^{(i)}+1} \mathbf{B}_\ell^{(i)}(\boldsymbol{\tau}_0)\bar{\mathbf{f}}_\ell^{(i)}, \tag{2.14}$$

$$\mathbf{S}^{(i)} = \sum_{\ell=0}^{L^{(i)}+1} \mathbf{B}_\ell^{(i)}(\boldsymbol{\tau}_0)\boldsymbol{\Phi}_\ell^{(i)}(n^{(i)})\mathbf{B}_\ell^{(i)\prime}(\boldsymbol{\tau}_0) + \sum_{\substack{\ell,m=0 \\ \ell \neq m}}^{L^{(i)}+1} \mathbf{B}_\ell^{(i)}(\boldsymbol{\tau}_0)\boldsymbol{\Phi}_{\ell m}^{(i)}(n^{(i)})\mathbf{B}_m^{(i)\prime}(\boldsymbol{\tau}_0), \tag{2.15}$$

where

$$\bar{\mathbf{f}}_\ell^{(i)} = \frac{1}{n^{(i)}}\sum_{j=1}^{n^{(i)}} \mathbf{f}_{\ell j}^{(i)} \quad \ell = 0, 1, \cdots, L^{(i)} + 1, \tag{2.16}$$

$$\boldsymbol{\Phi}_\ell^{(i)}(n^{(i)}) = \frac{1}{n^{(i)}-1}\sum_{j=1}^{n^{(i)}} (\mathbf{f}_{\ell j}^{(i)} - \bar{\mathbf{f}}_\ell^{(i)})(\mathbf{f}_{\ell j}^{(i)} - \bar{\mathbf{f}}_\ell^{(i)})' \quad \ell = 0, 1, \cdots, L^{(i)} + 1, \tag{2.17}$$

$$\boldsymbol{\Phi}_{\ell m}^{(i)}(n^{(i)}) = \frac{1}{n^{(i)}-1}\sum_{j=1}^{n^{(i)}} (\mathbf{f}_{\ell j}^{(i)} - \bar{\mathbf{f}}_\ell^{(i)})(\mathbf{f}_{mj}^{(i)} - \bar{\mathbf{f}}_m^{(i)})' \quad \ell \neq m = 0, \cdots, L^{(i)} + 1. \tag{2.18}$$

Thus, by Assumption 1 and the law of large numbers,

$$\mathrm{plim}_{n^{(i)}\to\infty}\bar{\mathbf{z}}^{(i)} = \boldsymbol{\mu}^{(i)}(\boldsymbol{\theta}_\infty), \quad \mathrm{plim}_{n^{(i)}\to\infty}\mathbf{S}^{(i)} = \boldsymbol{\Sigma}^{(i)}(\boldsymbol{\theta}_\infty), \tag{2.19}$$

where $\boldsymbol{\mu}^{(i)}(\boldsymbol{\theta})$ and $\boldsymbol{\Sigma}^{(i)}(\boldsymbol{\theta})$ are defined in (2.9) and (2.10). Hence, $\mathbf{c}$, defined in (2.8), satisfies $\mathrm{plim}_{n_m\to\infty}\mathbf{c} = \boldsymbol{\gamma}(\boldsymbol{\theta}_\infty)$, which implies $\mathrm{plim}_{n_m\to\infty}Q(\boldsymbol{\gamma}(\boldsymbol{\theta}_\infty),\ \mathbf{c}) = 0$ for Q defined in (2.7). Since $Q(\boldsymbol{\gamma}(\boldsymbol{\theta}),\ \mathbf{c}) \geq 0$ for all $\boldsymbol{\theta}$, and since $\hat{\boldsymbol{\theta}}$ minimizes Q, we have $\mathrm{plim}_{n_m\to\infty}Q(\boldsymbol{\gamma}(\hat{\boldsymbol{\theta}}),\ \mathbf{c}) = 0$. This and Assumption 2 imply

$$\mathrm{plim}_{n_m\to\infty}Q^{(i)}(\boldsymbol{\gamma}^{(i)}(\hat{\boldsymbol{\theta}}),\ \mathbf{c}^{(i)}) = \mathrm{plim}_{n_m\to\infty}[Q_1^{(i)}(\boldsymbol{\gamma}^{(i)}(\hat{\boldsymbol{\theta}}),\ \mathbf{c}^{(i)}) + Q_2^{(i)}(\boldsymbol{\gamma}^{(i)}(\hat{\boldsymbol{\theta}}),\ \mathbf{c}^{(i)})] = 0,$$

where

$$Q_1^{(i)}(\boldsymbol{\gamma}^{(i)}(\hat{\boldsymbol{\theta}}),\ \mathbf{c}^{(i)}) = \mathrm{tr}\{\mathbf{S}^{(i)}\boldsymbol{\Sigma}^{(i)-1}(\hat{\boldsymbol{\theta}})\} - \log|\mathbf{S}^{(i)}\boldsymbol{\Sigma}^{(i)-1}(\hat{\boldsymbol{\theta}})| - p^{(i)}$$

and

$$Q_2^{(i)}(\boldsymbol{\gamma}^{(i)}(\hat{\boldsymbol{\theta}}),\ \mathbf{c}^{(i)}) = [\bar{\mathbf{z}}^{(i)} - \boldsymbol{\mu}^{(i)}(\hat{\boldsymbol{\theta}})]'\boldsymbol{\Sigma}^{(i)-1}(\hat{\boldsymbol{\theta}})[\bar{\mathbf{z}}^{(i)} - \boldsymbol{\mu}^{(i)}(\hat{\boldsymbol{\theta}})].$$

Since $Q_1^{(i)}(\boldsymbol{\gamma}^{(i)}(\hat{\boldsymbol{\theta}}),\ \mathbf{c}^{(i)}) \xrightarrow{p} 0$ implies $\boldsymbol{\Sigma}^{(i)}(\hat{\boldsymbol{\theta}}) \xrightarrow{p} \boldsymbol{\Sigma}^{(i)}(\boldsymbol{\theta}_\infty)$ (see, e.g., Anderson (1987)), and since $Q_2^{(i)}(\boldsymbol{\gamma}^{(i)}(\hat{\boldsymbol{\theta}}),\ \mathbf{c}^{(i)}) \xrightarrow{p} 0$ implies that $\boldsymbol{\mu}^{(i)}(\hat{\boldsymbol{\theta}}) \xrightarrow{p} \boldsymbol{\mu}^{(i)}(\boldsymbol{\theta}_\infty)$, the result follows from Assumption 3. □

Hence, under very weak distributional specifications in Assumption 1, the estimator $\hat{\boldsymbol{\theta}}$ is consistent for the limiting true value $\boldsymbol{\theta}_\infty$. In fact, it is clear from the proof that the consistency of $\hat{\boldsymbol{\theta}}$ holds for any general mean and covariance structure model $\boldsymbol{\gamma}(\boldsymbol{\theta}) = E\{\mathbf{c}\}$ satisfying (2.19). To characterize the limiting behavior of $\hat{\boldsymbol{\theta}}$ in more detail, especially for the assessment of the so-called asymptotic robustness properties, it will be convenient to consider an expansion of $\hat{\boldsymbol{\theta}}$, not around the true value $\boldsymbol{\theta}_{\mathbf{n}}$ or the limiting true value $\boldsymbol{\theta}_\infty$, defined in (2.12) and (2.13), but around some other quantity $\bar{\boldsymbol{\theta}}_{\mathbf{n}}$ that depends on the unobservable sample moments of the latent variables $\mathbf{f}_{\ell j}^{(i)}$, $\ell = 1, 2, \cdots, L^{(i)} + 1$. For this purpose, let

$$\bar{\boldsymbol{\theta}}_{\mathbf{n}} = (\boldsymbol{\tau}_0', \ \boldsymbol{\rho}_{\mathbf{n}}', \ \boldsymbol{\varphi}_{\mathbf{n}}')', \tag{2.20}$$

where $\boldsymbol{\tau}_0$ and $\boldsymbol{\varphi}_{\mathbf{n}}$ are as in $\boldsymbol{\theta}_{\mathbf{n}}$,

$$\boldsymbol{\rho}_{\mathbf{n}} = (\boldsymbol{\rho}_{\mathbf{n}}^{(1)\prime}, \cdots, \boldsymbol{\rho}_{\mathbf{n}}^{(I)\prime})', \quad \boldsymbol{\rho}_{\mathbf{n}}^{(i)} = (\mathrm{v}'[\boldsymbol{\Phi}_1^{(i)}(n^{(i)})], \cdots, \mathrm{v}'[\boldsymbol{\Phi}_{L^{(i)}}^{(i)}(n^{(i)})])',$$

and $\boldsymbol{\Phi}_\ell^{(i)}(n^{(i)})$, $\ell = 1, 2, \cdots, L^{(i)}$ are defined in (2.17). Thus, in $\bar{\boldsymbol{\theta}}_{\mathbf{n}}$, the true value $\boldsymbol{\rho}_0$ in $\boldsymbol{\theta}_{\mathbf{n}}$ consisting of the true covariance matrices $\boldsymbol{\Phi}_{L^{(i)}}^{(i)0}$ of the random latent variables $\mathbf{f}_{\ell j}^{(i)}$, $\ell = 1, 2, \cdots, L^{(i)}$, is replaced by $\boldsymbol{\rho}_{\mathbf{n}}$ consisting of the unobservable sample moments $\boldsymbol{\Phi}_\ell^{(i)}(n^{(i)})$'s. While statistical inference is to be made for $\boldsymbol{\theta}_{\mathbf{n}}$ in (2.12), $\bar{\boldsymbol{\theta}}_{\mathbf{n}}$ with an artificial quantity $\boldsymbol{\rho}_{\mathbf{n}}$ plays a useful role in assessing the property of $\hat{\boldsymbol{\tau}}$ in $\hat{\boldsymbol{\theta}}$, as well as in characterizing the limiting distribution of the whole $\hat{\boldsymbol{\theta}}$ without specifying any moments for $\mathbf{f}_{\ell j}^{(i)}$ higher than the second order. To obtain an expansion of $\hat{\boldsymbol{\theta}}$ around $\bar{\boldsymbol{\theta}}_{\mathbf{n}}$, we need some smoothness conditions for $\mathbf{b}^{(i)}(\boldsymbol{\tau})$, $\mathbf{B}_\ell^{(i)}(\boldsymbol{\tau})$, and $\boldsymbol{\Phi}_0^{(i)}(\boldsymbol{\tau})$, and the full column rank of $\mathbf{F}_\infty = \mathbf{F}(\boldsymbol{\theta}_\infty)$, where

$$\mathbf{F}(\boldsymbol{\theta}) = [\mathbf{F}^{(1)\prime}(\boldsymbol{\theta}), \cdots, \mathbf{F}^{(I)\prime}(\boldsymbol{\theta})]', \quad \mathbf{F}^{(i)}(\boldsymbol{\theta}) = \frac{\partial \boldsymbol{\gamma}^{(i)}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'}. \tag{2.21}$$

Since the linear independency of the columns of $\mathbf{F}$ associated with the $\boldsymbol{\rho}$ and $\boldsymbol{\varphi}$ parts of $\boldsymbol{\theta}$ is trivial, we need to assume only that the $\boldsymbol{\tau}$ part of the model is specified without any redundancy:

Assumption 4 *For all $i = 1, 2, \cdots, I$, $\mathbf{b}^{(i)}(\boldsymbol{\tau})$, $\mathbf{B}_\ell^{(i)}(\boldsymbol{\tau})$, and $\boldsymbol{\Phi}_0^{(i)}(\boldsymbol{\tau})$ are twice continuously differentiable in $\boldsymbol{\Gamma}_{\boldsymbol{\tau}}$. The first $d_{\boldsymbol{\tau}}$ columns of $\mathbf{F}_\infty$ in (2.21) are linearly independent.*

We use the notation $\mathbf{D}_p$ to denote the $p^2 \times p(p+1)/2$ matrix with 0's and 1's satisfying $\mathrm{vec}(\mathbf{A}) = \mathbf{D}_p \mathrm{v}(\mathbf{A})$ for any $p \times p$ symmetric matrix $\mathbf{A}$, where $\mathrm{vec}\mathbf{A}$ is the

$p^2 \times 1$ vector listing the p columns of $\mathbf{A}$ in one vector starting with the first column. For more details, see, e.g., Fuller (1987) and Magnus and Neuedecker (1988). Also, let $\mathbf{D}_p^+$ denote the Moore-Penrose generalized inverse of $\mathbf{D}_p$, i.e., $\mathbf{D}_p^+ = (\mathbf{D}_p'\mathbf{D}_p)^{-1}\mathbf{D}_p'$. Note that if $\mathbf{W} \sim \mathbf{W}_p(\Sigma,\ d)$ (Wishart distribution), then as $d \to \infty$,

$$\sqrt{d}\ \mathrm{v}(d^{-1}\mathbf{W} - \Sigma) \xrightarrow{L} \mathbf{N}(\mathbf{0}, \Gamma), \tag{2.22}$$

$$\Gamma = 2\mathbf{D}_p^+(\Sigma \otimes \Sigma)\mathbf{D}_p^{+\prime}, \quad \Gamma^{-1} = \frac{1}{2}\mathbf{D}_p'(\Sigma^{-1} \otimes \Sigma^{-1})\mathbf{D}_p. \tag{2.23}$$

See, e.g., Fuller (1987, p. 386). The next theorem gives an expansion of $\hat{\boldsymbol{\theta}}$ around $\bar{\boldsymbol{\theta}}_{\mathbf{n}}$.

Theorem 2 *Assume that model (2.3) holds. Let Assumptions 1-4 hold. Then,*

$$\hat{\boldsymbol{\theta}} - \bar{\boldsymbol{\theta}}_{\mathbf{n}} = (\mathbf{F}_\infty'\Omega_\infty^{-1}\mathbf{F}_\infty)^{-1}\mathbf{F}_\infty'\Omega_\infty^{-1}[\mathbf{c} - \gamma(\bar{\boldsymbol{\theta}}_{\mathbf{n}})] + o_p(n_m^{-1/2}), \tag{2.24}$$

where $\bar{\boldsymbol{\theta}}_{\mathbf{n}}$, $\mathbf{F}_\infty = \mathbf{F}(\boldsymbol{\theta}_\infty)$, *and* $\mathbf{D}_{p^{(i)}}$ *are defined in (2.20), (2.21), and (2.23), respectively,* $\Omega_\infty^{-1} = \Omega^{-1}(\boldsymbol{\theta}_\infty)$,

$$\Omega^{-1}(\boldsymbol{\theta}) = blockdiag\{r^{(1)}\Omega^{(1)-1}(\boldsymbol{\theta}), \cdots, r^{(I)}\Omega^{(I)-1}(\boldsymbol{\theta})\}, \tag{2.25}$$

$$\Omega^{(i)-1}(\boldsymbol{\theta}) = \begin{pmatrix} \Sigma^{(i)-1}(\boldsymbol{\theta}) & \mathbf{0} \\ \mathbf{0} & \Gamma^{(i)-1}(\boldsymbol{\theta}) \end{pmatrix},$$

$$\Gamma^{(i)-1}(\boldsymbol{\theta}) = \frac{1}{2}\mathbf{D}_{p^{(i)}}'[\Sigma^{(i)-1}(\boldsymbol{\theta}) \otimes \Sigma^{(i)-1}(\boldsymbol{\theta})]\mathbf{D}_{p^{(i)}}.$$

Proof: By (2.9), (2.10), (2.14), (2.15), and (2.20) $\mathbf{c} - \gamma(\bar{\boldsymbol{\theta}}_{\mathbf{n}})$ consists of

$$\begin{aligned}
\bar{\mathbf{z}}^{(i)} - \boldsymbol{\mu}^{(i)}(\bar{\boldsymbol{\theta}}_{\mathbf{n}}) &= \sum_{\ell=0}^{L^{(i)}} \mathbf{B}_\ell^{(i)}(\boldsymbol{\tau}_0)\bar{\mathbf{f}}_\ell^{(i)}, \\
\mathrm{v}\{\mathbf{S}^{(i)} - \Sigma^{(i)}(\bar{\boldsymbol{\theta}}_{\mathbf{n}})\} &= \mathbf{D}_{p^{(i)}}^+[\mathbf{B}_0^{(i)}(\boldsymbol{\tau}_0) \otimes \mathbf{B}_0^{(i)}(\boldsymbol{\tau}_0)]\mathbf{D}_{q_0^{(i)}}\mathrm{v}[\Phi_0^{(i)}(n^{(i)}) - \Phi_0^{(i)}(\boldsymbol{\tau}_0)] \\
&\quad + 2\sum_{\substack{\ell,m=0 \\ \ell<m}}^{L^{(i)}+1} \mathbf{D}_{p^{(i)}}^+[\mathbf{B}_m^{(i)}(\boldsymbol{\tau}_0) \otimes \mathbf{B}_\ell^{(i)}(\boldsymbol{\tau}_0)]\mathrm{vec}[\Phi_{\ell m}^{(i)}(n^{(i)})]
\end{aligned} \tag{2.26}$$

where $\bar{\mathbf{f}}_\ell^{(i)}$, $\mathbf{\Phi}_0^{(i)}(n^{(i)})$, and $\mathbf{\Phi}_{\ell m}^{(i)}(n^{(i)})$ are defined in (2.16), (2.17), and (2.18). By Assumption 1 *i-iii*), $\mathbf{f}_{0j}^{(i)}$'s are normally distributed and $\mathbf{\Phi}_0^{(i)}(n^{(i)}) - \mathbf{\Phi}_0^{(i)}(\boldsymbol{\tau}_0) = O_p(n_m^{-1/2})$. By the independence and the existence of the second moments of $\mathbf{f}_{\ell j}^{(i)}$, $\ell = 1, 2, \cdots, L^{(i)}$, in Assumption 1 *i*), and by the limiting condition for $\mathbf{f}_{L^{(i)}+1,j}^{(i)}$ in Assumption 1 *iv*), we have $\mathbf{\Phi}_{\ell m}^{(i)}(n^{(i)}) = O_p(n_m^{-1/2})$, $\ell \neq m$. Thus,

$$\mathbf{c} - \boldsymbol{\gamma}(\bar{\boldsymbol{\theta}}_{\mathbf{n}}) = O_p(n_m^{-1/2}). \tag{2.27}$$

Note that (2.26) does not contain $\mathbf{\Phi}_\ell^{(i)}(n^{(i)})$, $\ell > 0$. This is due to the use of $\bar{\boldsymbol{\theta}}_{\mathbf{n}}$ instead of $\boldsymbol{\theta}_{\mathbf{n}}$ in the expansion. We can write

$$\left.\frac{\partial Q(\boldsymbol{\gamma}(\boldsymbol{\theta}),\ \mathbf{c})}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = \left.\frac{\partial Q(\boldsymbol{\gamma}(\boldsymbol{\theta}),\ \mathbf{c})}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\bar{\boldsymbol{\theta}}_{\mathbf{n}}} + \left.\frac{\partial^2 Q(\boldsymbol{\gamma}(\boldsymbol{\theta}),\ \mathbf{c})}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}(\hat{\boldsymbol{\theta}} - \bar{\boldsymbol{\theta}}_{\mathbf{n}}), \tag{2.28}$$

where $\boldsymbol{\theta}^*$ is on the line segment between $\hat{\boldsymbol{\theta}}$ and $\bar{\boldsymbol{\theta}}_{\mathbf{n}}$. Since $\hat{\boldsymbol{\theta}} \xrightarrow{p} \boldsymbol{\theta}_\infty$ (Theorem 1) and $\bar{\boldsymbol{\theta}}_{\mathbf{n}} \xrightarrow{p} \boldsymbol{\theta}_\infty$ (Assumption 1), it follows that $\boldsymbol{\theta}^* \xrightarrow{p} \boldsymbol{\theta}_\infty$. By Assumption 1 and the definition of $\mathbf{\Gamma}_{\boldsymbol{\theta}}$, $\boldsymbol{\theta}_\infty$ is an interior point of $\mathbf{\Gamma}_{\boldsymbol{\theta}}$. Thus, with probability approaching one, $\hat{\boldsymbol{\theta}}$ is well defined and the left-hand side of (2.28) is 0. For the j-th element θ_j of $\boldsymbol{\theta}$,

$$\begin{aligned}
\frac{\partial Q^{(i)}(\boldsymbol{\gamma}^{(i)}(\boldsymbol{\theta}),\ \mathbf{c}^{(i)})}{\partial \theta_j} &= \mathrm{tr}\{\mathbf{\Sigma}^{(i)-1}(\boldsymbol{\theta})[\mathbf{\Sigma}^{(i)}(\boldsymbol{\theta}) - \mathbf{S}^{(i)}]\mathbf{\Sigma}^{(i)-1}(\boldsymbol{\theta})\frac{\partial \mathbf{\Sigma}^{(i)}(\boldsymbol{\theta})}{\theta_j}\} \\
&\quad -2\frac{\partial \boldsymbol{\mu}^{(i)\prime}(\boldsymbol{\theta})}{\partial \theta_j}\mathbf{\Sigma}^{(i)-1}(\boldsymbol{\theta})[\bar{\mathbf{z}}^{(i)} - \boldsymbol{\mu}^{(i)}(\boldsymbol{\theta})] \\
&\quad -[\bar{\mathbf{z}}^{(i)} - \boldsymbol{\mu}^{(i)}(\boldsymbol{\theta})]'\mathbf{\Sigma}^{(i)-1}(\boldsymbol{\theta})\frac{\partial \mathbf{\Sigma}^{(i)}(\boldsymbol{\theta})}{\theta_j}\mathbf{\Sigma}^{(i)-1}(\boldsymbol{\theta})[\bar{\mathbf{z}}^{(i)} - \boldsymbol{\mu}^{(i)}(\boldsymbol{\theta})].
\end{aligned}$$

It follows that

$$\begin{aligned}
\left.\frac{\partial Q(\boldsymbol{\gamma}(\boldsymbol{\theta}),\ \mathbf{c})}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\bar{\boldsymbol{\theta}}_{\mathbf{n}}} &= -2\sum_{i=1}^{I}\frac{n^{(i)}}{n}\{\frac{\partial \mathrm{v}'[\mathbf{\Sigma}^{(i)}(\bar{\boldsymbol{\theta}}_{\mathbf{n}})]}{\partial \boldsymbol{\theta}}\mathbf{\Gamma}^{(i)-1}(\bar{\boldsymbol{\theta}}_{\mathbf{n}}) \\
&\quad \times \mathrm{v}[\mathbf{S}^{(i)} - \mathbf{\Sigma}^{(i)}(\bar{\boldsymbol{\theta}}_{\mathbf{n}})] + \frac{\partial \boldsymbol{\mu}^{(i)\prime}(\bar{\boldsymbol{\theta}}_{\mathbf{n}})}{\partial \boldsymbol{\theta}}\mathbf{\Sigma}^{(i)-1}(\bar{\boldsymbol{\theta}}_{\mathbf{n}})
\end{aligned}$$

$$\times[\bar{\mathbf{z}}^{(i)} - \boldsymbol{\mu}^{(i)}(\bar{\boldsymbol{\theta}}_{\mathbf{n}})]\} + O_p(n_m^{-1})$$
$$= -2\mathbf{F}'_{\infty}\boldsymbol{\Omega}_{\infty}^{-1}[\mathbf{c} - \boldsymbol{\gamma}(\bar{\boldsymbol{\theta}}_{\mathbf{n}})] + o_p(n_m^{-1/2}). \quad (2.29)$$

Similarly, we have

$$\begin{aligned} \mathrm{plim}_{n_m\to\infty} \frac{\partial^2 Q^{(i)}(\boldsymbol{\gamma}^{(i)}(\boldsymbol{\theta}),\ \mathbf{c}^{(i)})}{\partial\theta_j\partial\theta_k}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} &= \mathrm{tr}\{\boldsymbol{\Sigma}^{(i)-1}(\boldsymbol{\theta}_{\infty})\frac{\partial\boldsymbol{\Sigma}^{(i)}(\boldsymbol{\theta}_{\infty})}{\partial\theta_k} \\ &\quad \times\boldsymbol{\Sigma}^{(i)-1}(\boldsymbol{\theta}_{\infty})\frac{\partial\boldsymbol{\Sigma}^{(i)}(\boldsymbol{\theta}_{\infty})}{\partial\theta_j}\} \\ &\quad +2\frac{\partial\boldsymbol{\mu}^{(i)\prime}(\boldsymbol{\theta}_{\infty})}{\partial\theta_j}\boldsymbol{\Sigma}^{(i)-1}(\boldsymbol{\theta}_{\infty})\frac{\partial\boldsymbol{\mu}^{(i)}(\boldsymbol{\theta}_{\infty})}{\partial\theta_k}, \end{aligned}$$

and

$$\mathrm{plim}_{n_m\to\infty} \frac{\partial^2 Q(\boldsymbol{\gamma}(\boldsymbol{\theta}),\ \mathbf{c})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} = 2\mathbf{F}'_{\infty}\boldsymbol{\Omega}_{\infty}^{-1}\mathbf{F}_{\infty}. \quad (2.30)$$

Thus, the result follows from (2.28), (2.29), and (2.30). □

Theorem 2 expresses the leading term of $\hat{\boldsymbol{\theta}} - \bar{\boldsymbol{\theta}}_{\mathbf{n}}$ in terms of $\mathbf{c} - \boldsymbol{\gamma}(\bar{\boldsymbol{\theta}}_{\mathbf{n}})$. Note that the use of $\bar{\boldsymbol{\theta}}_{\mathbf{n}}$ in Theorem 2 produced an expansion of $\hat{\boldsymbol{\theta}}$ around $\bar{\boldsymbol{\theta}}_{\mathbf{n}}$ with the existence of only second moments of $\mathbf{f}_{\ell j}^{(i)}$, $\ell = 1, \cdots, L^{(i)}$. It can be shown from the proof that the expansion in Theorem 2 holds for the general model $\boldsymbol{\gamma}(\boldsymbol{\theta}) = E\{\mathbf{c}\}$ and for any $\bar{\boldsymbol{\theta}}_{\mathbf{n}}$ with $\bar{\boldsymbol{\theta}}_{\mathbf{n}} \xrightarrow{p} \boldsymbol{\theta}_{\infty}$ provided (2.27) holds. But, the special choice of $\bar{\boldsymbol{\theta}}_{\mathbf{n}}$ in (2.20) for model (2.3) makes the result of Theorem 2 practically meaningful. Note that for $\bar{\boldsymbol{\theta}}_{\mathbf{n}}$ in (2.20) the $\boldsymbol{\tau}$-part of $\hat{\boldsymbol{\theta}} - \bar{\boldsymbol{\theta}}_{\mathbf{n}}$ is $\hat{\boldsymbol{\tau}} - \boldsymbol{\tau}_0$ and the $\boldsymbol{\varphi}$-part of $\hat{\boldsymbol{\theta}} - \bar{\boldsymbol{\theta}}_{\mathbf{n}}$ is $\hat{\boldsymbol{\varphi}} - \boldsymbol{\varphi}_{\mathbf{n}}$. Thus, Theorem 2 in fact produces the expansion relevant for assessing the limiting distribution of $\sqrt{n}[(\hat{\boldsymbol{\tau}}',\ \hat{\boldsymbol{\varphi}}')' - (\boldsymbol{\tau}_0',\ \boldsymbol{\varphi}_{\mathbf{n}}')']$, which in turn shows the wide applicability of the large sample estimated covariance matrix of $\hat{\boldsymbol{\tau}}$ derived under the normal-independence model. Here, the normal-independence model refers to the model assumed in the multi-sample option of the existing computer packages where observations, i.e., all underlying latent variables, are assumed to be normally distributed and the samples are assumed to be independent. The function Q in (2.11) corresponds to the likelihood for the normal-independence model. If model (2.3) is the true model, the normal-independence model is an incorrect model used only for computation. Since the existing computer packages compute the large sample standard error or estimated covariance matrix of the estimates, it is useful to consider their applicability and required adjustments for model (2.3). If all $\mathbf{f}_{\ell j}^{(i)}$, $\ell = 0, 1, \cdots, L^{(i)} + 1$, in model (2.3) were incorrectly assumed normal, then we would be estimating

$$\boldsymbol{\theta}_0 = (\boldsymbol{\tau}_0', \; \boldsymbol{\rho}_0', \; \boldsymbol{\varphi}_0')', \tag{2.31}$$

where $\boldsymbol{\varphi}_0$ would consist of the true means and covariance matrices of the incorrectly assumed normal $\mathbf{f}_{L^{(i)}+1,j}^{(i)}$'s. Under the normal-independence model, the limiting distribution of the parameter vector $\hat{\boldsymbol{\theta}}$ would be

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{L} \mathbf{N}(\mathbf{0}, \; \mathbf{V}_N(\boldsymbol{\theta}_0)), \quad \text{as} \;\; n_m \to \infty, \tag{2.32}$$

where

$$\mathbf{V}_N(\boldsymbol{\theta}) = [\mathbf{F}'(\boldsymbol{\theta})\boldsymbol{\Omega}^{-1}(\boldsymbol{\theta})\mathbf{F}(\boldsymbol{\theta})]^{-1}. \tag{2.33}$$

The matrices $\mathbf{F}(\boldsymbol{\theta})$ and $\boldsymbol{\Omega}^{-1}(\boldsymbol{\theta})$ are defined in (2.21) and (2.25), respectively. This follows from the standard normal likelihood theory, or from (2.22) and (2.24) for model (2.3) with no fixed or unspecified random latent variables, i.e., with no $\mathbf{f}_{\ell j}^{(i)}$, $\ell > 0$. Thus, the computer packages based on the normal-independence model would compute $n^{-1}\mathbf{V}_N(\hat{\boldsymbol{\theta}})$ as the estimated large sample covariance matrix of $\hat{\boldsymbol{\theta}}$. Under the normal-independence model, $\mathbf{V}_N(\hat{\boldsymbol{\theta}}) \xrightarrow{P} \mathbf{V}_N(\boldsymbol{\theta}_0)$. Under our model (2.3) with Assumption 1, we do not expect that $n^{-1}\mathbf{V}_N(\hat{\boldsymbol{\theta}})$ can be correctly used for inferences, such as asymptotic confidence intervals and testing, concerning all the elements of $\boldsymbol{\theta}_{\mathbf{n}}$ in (2.12). However, the next theorem shows, based on the expansion in Theorem 2, that $n^{-1}\mathbf{V}_N(\hat{\boldsymbol{\theta}})$ can still be used correctly to make inferences for $\boldsymbol{\tau}$ and requires only a small adjustment for inferences for $\boldsymbol{\varphi}_{\mathbf{n}}$. For this, we write $\mathbf{V}_N(\boldsymbol{\theta})$ defined in (2.33) as

$$\mathbf{V}_N(\boldsymbol{\theta}) = \begin{pmatrix} \mathbf{V}_{N\boldsymbol{\tau}\boldsymbol{\tau}}(\boldsymbol{\theta}) & \mathbf{V}_{N\boldsymbol{\tau}\boldsymbol{\rho}}(\boldsymbol{\theta}) & \mathbf{V}_{N\boldsymbol{\tau}\boldsymbol{\varphi}}(\boldsymbol{\theta}) \\ \mathbf{V}_{N\boldsymbol{\rho}\boldsymbol{\tau}}(\boldsymbol{\theta}) & \mathbf{V}_{N\boldsymbol{\rho}\boldsymbol{\rho}}(\boldsymbol{\theta}) & \mathbf{V}_{N\boldsymbol{\rho}\boldsymbol{\varphi}}(\boldsymbol{\theta}) \\ \mathbf{V}_{N\boldsymbol{\varphi}\boldsymbol{\tau}}(\boldsymbol{\theta}) & \mathbf{V}_{N\boldsymbol{\varphi}\boldsymbol{\rho}}(\boldsymbol{\theta}) & \mathbf{V}_{N\boldsymbol{\varphi}\boldsymbol{\varphi}}(\boldsymbol{\theta}) \end{pmatrix}, \tag{2.34}$$

where $\mathbf{V}_{N\boldsymbol{\tau}\boldsymbol{\tau}}(\boldsymbol{\theta})$ is $d_{\boldsymbol{\tau}} \times d_{\boldsymbol{\tau}}$, $\mathbf{V}_{N\boldsymbol{\rho}\boldsymbol{\rho}}(\boldsymbol{\theta})$ is $d_{\boldsymbol{\rho}} \times d_{\boldsymbol{\rho}}$, and $\mathbf{V}_{N\boldsymbol{\varphi}\boldsymbol{\varphi}}(\boldsymbol{\theta})$ is $d_{\boldsymbol{\varphi}} \times d_{\boldsymbol{\varphi}}$.

Theorem 3 *Let model (2.3) and Assumptions 1-4 hold. Then as $n_m \to \infty$,*

$$\sqrt{n}\left[\begin{pmatrix} \hat{\boldsymbol{\tau}} \\ \hat{\boldsymbol{\varphi}} \end{pmatrix} - \begin{pmatrix} \boldsymbol{\tau}_0 \\ \boldsymbol{\varphi}_{\mathbf{n}} \end{pmatrix}\right] \xrightarrow{L} \mathbf{N}(\mathbf{0}, \mathbf{V}^*(\boldsymbol{\theta}_\infty)),$$

where

$$\mathbf{V}^*(\boldsymbol{\theta}) = \begin{pmatrix} \mathbf{V}_{N\boldsymbol{\tau}\boldsymbol{\tau}}(\boldsymbol{\theta}) & \mathbf{V}_{N\boldsymbol{\tau}\boldsymbol{\varphi}}(\boldsymbol{\theta}) \\ \mathbf{V}_{N\boldsymbol{\varphi}\boldsymbol{\tau}}(\boldsymbol{\theta}) & \mathbf{V}_{N\boldsymbol{\varphi}\boldsymbol{\varphi}}(\boldsymbol{\theta}) - \mathbf{U}(\boldsymbol{\theta}) \end{pmatrix},$$

$$\mathbf{U}(\boldsymbol{\theta}) = blockdiag\{\frac{1}{\tau^{(1)}}\mathbf{U}^{(1)}(\boldsymbol{\varphi}), \cdots, \frac{1}{\tau^{(I)}}\mathbf{U}^{(I)}(\boldsymbol{\varphi})\},$$

$$\mathbf{U}^{(i)}(\boldsymbol{\varphi}) = \begin{pmatrix} \boldsymbol{\Phi}^{(i)}_{L^{(i)}+1} & \mathbf{0} \\ \mathbf{0} & 2\mathbf{D}^{+}_{q^{(i)}_{L^{(i)}+1}}(\boldsymbol{\Phi}^{(i)}_{L^{(i)}+1} \otimes \boldsymbol{\Phi}^{(i)}_{L^{(i)}+1})\mathbf{D}^{+\prime}_{q^{(i)}_{L^{(i)}+1}} \end{pmatrix},$$

and $\boldsymbol{\Phi}^{(i)}_{L^{(i)}+1}$ is defined in Assumption 1-iv). In addition,

$$\mathbf{V}^*(\hat{\boldsymbol{\theta}}) \xrightarrow{p} \mathbf{V}^*(\boldsymbol{\theta}_\infty).$$

Proof: Note that the quantities in (2.26) are functions of the unobservable sample averages, the sample covariance matrices of normal $\mathbf{f}^{(i)}_{0j}$, and the sample cross-products of independent $\mathbf{f}_{\ell j}^{(i)}$ and $\mathbf{f}_{mj}^{(i)}$ $\ell \neq m$. Under Assumption 1, these quantities multiplied by $\sqrt{n}$ converge jointly to a normal distribution with zero mean. Note that $\mathbf{c}-\boldsymbol{\gamma}(\bar{\boldsymbol{\theta}}_{\mathbf{n}})$ consists of the quantities in (2.26). By Theorem 2, the leading term in $\hat{\boldsymbol{\theta}} - \bar{\boldsymbol{\theta}}_{\mathbf{n}}$ is $\mathbf{c} - \boldsymbol{\gamma}(\bar{\boldsymbol{\theta}}_{\mathbf{n}})$ multiplied by a coefficient matrix evaluated at $\boldsymbol{\theta}_\infty$. It is clear that under the normal-independence model the leading term of $\hat{\boldsymbol{\theta}} - \bar{\boldsymbol{\theta}}_{\mathbf{n}}$ has the same form with the coefficient matrix evaluated at $\boldsymbol{\theta}_0$. Hence, the limiting distribution of $\sqrt{n}(\hat{\boldsymbol{\theta}}-\bar{\boldsymbol{\theta}}_{\mathbf{n}})$ under Assumption 1 is the same as that under the normal-independence model with $\boldsymbol{\theta}_0$ replaced by $\boldsymbol{\theta}_\infty$. This gives the limiting distribution for $\sqrt{n}(\hat{\boldsymbol{\tau}} - \boldsymbol{\tau}_0)$, since the $\boldsymbol{\tau}$-part of $\hat{\boldsymbol{\theta}} - \bar{\boldsymbol{\theta}}_{\mathbf{n}}$ is $\hat{\boldsymbol{\tau}} - \boldsymbol{\tau}_0$. Also, the $\boldsymbol{\varphi}$-part of $\hat{\boldsymbol{\theta}} - \bar{\boldsymbol{\theta}}_{\mathbf{n}}$ is $\hat{\boldsymbol{\varphi}} - \boldsymbol{\varphi}_{\mathbf{n}}$ and the $\boldsymbol{\varphi}$-part of $\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0$ in (2.32) is $\hat{\boldsymbol{\varphi}} - \boldsymbol{\varphi}_0$, where $\boldsymbol{\varphi}_0$ consists of the true means and covariance matrices of $\mathbf{f}_{L^{(i)}+1,j}^{(i)}$ believed to be normally distributed. Write

$$\boldsymbol{\alpha}_{1\mathbf{n}} = \sqrt{n}\begin{pmatrix} \hat{\boldsymbol{\tau}} - \boldsymbol{\tau}_0 \\ \hat{\boldsymbol{\varphi}} - \boldsymbol{\varphi}_0 \end{pmatrix} = \sqrt{n}\begin{pmatrix} \hat{\boldsymbol{\tau}} - \boldsymbol{\tau}_0 \\ \hat{\boldsymbol{\varphi}} - \boldsymbol{\varphi}_{\mathbf{n}} \end{pmatrix} + \sqrt{n}\begin{pmatrix} \mathbf{0} \\ \boldsymbol{\varphi}_{\mathbf{n}} - \boldsymbol{\varphi}_0 \end{pmatrix} = \boldsymbol{\alpha}_{2\mathbf{n}} + \boldsymbol{\alpha}_{3\mathbf{n}}.$$

The above discussion showed that $\boldsymbol{\alpha}_{2\mathbf{n}} \xrightarrow{L} \mathrm{N}(\mathbf{0},\ \mathbf{V}^*(\boldsymbol{\theta}_\infty))$ under model (2.3) and $\boldsymbol{\alpha}_{2\mathbf{n}} \xrightarrow{L} \mathrm{N}(\mathbf{0},\ \mathbf{V}^*(\boldsymbol{\theta}_0))$ under the normal-independence model. By (2.32), under the normal-independence model, $\boldsymbol{\alpha}_{1\mathbf{n}} \xrightarrow{L} \mathrm{N}(\mathbf{0},\ \mathbf{V}_1(\boldsymbol{\theta}_0))$, where

$$\mathbf{V}_1(\boldsymbol{\theta}) = \begin{pmatrix} \mathbf{V}_{N\boldsymbol{\tau}\boldsymbol{\tau}}(\boldsymbol{\theta}) & \mathbf{V}_{N\boldsymbol{\tau}\boldsymbol{\varphi}}(\boldsymbol{\theta}) \\ \mathbf{V}_{N\boldsymbol{\varphi}\boldsymbol{\tau}}(\boldsymbol{\theta}) & \mathbf{V}_{N\boldsymbol{\varphi}\boldsymbol{\varphi}}(\boldsymbol{\theta}) \end{pmatrix}.$$

Since $\boldsymbol{\varphi}_{\mathbf{n}}$ consists of $\bar{\mathbf{f}}_{L^{(i)}+1}^{(i)}$ and $\boldsymbol{\Phi}_{L^{(i)}+1}^{(i)}(n^{(i)})$, under the normal-independence model, $\boldsymbol{\alpha}_{3\mathbf{n}} \xrightarrow{L} \mathrm{N}(\mathbf{0},\ \mathbf{V}_3(\boldsymbol{\theta}_0))$, where

$$\mathbf{V}_3(\boldsymbol{\theta}) = \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{U}(\boldsymbol{\theta}) \end{pmatrix}.$$

Since the leading term of $\boldsymbol{\alpha}_{2n}$ is a function of the quantities in (2.26), $\boldsymbol{\alpha}_{2n}$ and $\boldsymbol{\alpha}_{3n}$ are uncorrelated in the joint limiting distribution. Hence, as functions of $\boldsymbol{\theta}$,

$$\mathbf{V}_1(\boldsymbol{\theta}) = \mathbf{V}^*(\boldsymbol{\theta}) + \mathbf{V}_3(\boldsymbol{\theta}),$$

for any $\boldsymbol{\theta}$. Thus, $\mathbf{V}^*(\boldsymbol{\theta}_\infty) = \mathbf{V}_1(\boldsymbol{\theta}_\infty) - \mathbf{V}_3(\boldsymbol{\theta}_\infty)$, and the limiting distribution result under model (2.3) follows. The consistency of the estimated covariance matrix follows from the convergence in probability of $\hat{\boldsymbol{\theta}}$ to $\boldsymbol{\theta}_\infty$. □

Theorem 3 shows the usefulness of $\hat{\boldsymbol{\tau}}$ obtained by minimizing (2.11) or computed by the existing computer packages for the general latent variable model (2.3) with the errors-in-variables parameterization specified in Assumption 1. Theorem 3 states that if we blindly assume normality for the observations, and if the normality-based procedure was used to obtain $\hat{\boldsymbol{\tau}}$ and its estimated large sample covariance matrix, then the resulting statistical inferences for $\boldsymbol{\tau}$ are valid for model (2.3) with Assumption 1. Note that the distribution of the observations $\mathbf{z}_j^{(i)}$ in model (2.3) with Assumption 1 belongs to a broad class of non-normal and unspecified distributions. As noted in Section 2.3, in the parameterization of Assumption 1, the parameter $\boldsymbol{\tau}$ contains information concerning the means and the relationships among variables that may be common over populations or may be contrasting populations, and the variance-covariance parameters restricted over populations such as measurement error variances. Hence, the inferences for $\boldsymbol{\tau}$ cover many of the questions relevant for multi-population analysis. Thus, Theorem 3 showed the validity and usefulness of the multi-sample analysis using the normality-based packages, provided that the parameterization and modeling of (2.3) and Assumption 1 are used.

Another aspect of the statistical analysis under model (2.3) that may be of practical interest is to make inferences concerning the fixed latent vectors $\mathbf{f}_{L^{(i)}+1,j}^{(i)}$. As in (2.9), the mean for the i-th population is a function of $\boldsymbol{\tau}$ and the unobserved averages $\bar{\mathbf{f}}^{(i)}_{L^{(i)}+1}$. As mentioned in Section 2, the fixed latent vectors may represent aspects of the populations not explained by random sampling or independence among populations, and may play an important role in characterizing the relationships and differences among sampled individuals from I groups. Thus, inferences for $\boldsymbol{\varphi}_{\mathbf{n}}$ consisting of $\bar{\mathbf{f}}^{(i)}_{L^{(i)}+1}$ and $\bar{\boldsymbol{\Phi}}^{(i)}_{L^{(i)}+1}(n^{(i)})$ and for functions of $\boldsymbol{\varphi}_{\mathbf{n}}$ and $\boldsymbol{\tau}$ may be of interest. Theorem 3 showed that the asymptotic covariance matrix estimate for $\hat{\boldsymbol{\varphi}}$ from the normality-based packages is not correct for the general model (2.3), but that an asymptotically correct covariance matrix estimate can be obtained by simply subtracting $n^{-1}\mathbf{U}(\hat{\boldsymbol{\theta}})$. Given $\hat{\boldsymbol{\varphi}}$, the computation of $\mathbf{U}(\hat{\boldsymbol{\theta}})$ is immediate. Hence, with a trivial adjustment, the asymptotically correct inferences for quantities involving $\boldsymbol{\varphi}$ for model (2.3) can be carried out using $\hat{\boldsymbol{\theta}}$ and its estimated covariance matrix obtained by the normality-based packages. This is true for any distributions of random latent vectors $\mathbf{f}^{(i)}_{\ell j}$, $\ell = 1, 2, \cdots, L^{(i)}$, including discrete distributions and those without any moments higher than two.

To obtain the limiting distribution of $\hat{\boldsymbol{\rho}} - \boldsymbol{\rho}_0$ with $\boldsymbol{\rho}_0$ containing the true covariance matrices of the random latent vectors $\mathbf{f}^{(i)}_{\ell j}$, $\ell = 1, 2, \cdots, L^{(i)}$, with unspecified distributions, we need to assume the existence of the fourth moments of $\mathbf{f}^{(i)}_{\ell j}$, $\ell = 1, 2, \cdots, L^{(i)}$, in model (2.3). Since the form of the third and fourth moments depend on the distribution, the asymptotic covariance matrix of $\hat{\boldsymbol{\rho}}$ obtained under the normal-independence model is not generally appropriate for use in making inference for $\boldsymbol{\rho}_0$ under model (2.3) with non-normal random latent vectors.

The following theorem gives the limiting distribution for the whole $\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_{\mathbf{n}}$, assuming finite fourth moments for the non-normal latent variables, $\mathbf{f}^{(i)}_{\ell j}$, $\ell = 1, 2, \cdots, L^{(i)}$.

Theorem 4 *Let model (2.3) and Assumptions 1-4 hold. In addition, assume that the non-normal random latent variables* $\mathbf{f}_{\ell j}^{(i)}$, $\ell = 1, 2, \cdots, L^{(i)}$; $i = 1, 2, \cdots, I$, *have finite fourth moments. Then, as* $n_m \to \infty$,

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_{\mathbf{n}}) \xrightarrow{L} \mathbf{N}(\mathbf{0}, \mathbf{V}_G),$$

where

$$\mathbf{V}_G = \mathbf{V}_N(\boldsymbol{\theta}_\infty) + \begin{pmatrix} \mathbf{0} & \mathbf{G}_1 & \mathbf{0} \\ \mathbf{G}_1' & \boldsymbol{\Delta} + \mathbf{G}_2 + \mathbf{G}_2' & \mathbf{G}_3' \\ \mathbf{0} & \mathbf{G}_3 & -\mathbf{U}(\boldsymbol{\theta}_\infty) \end{pmatrix},$$

$\mathbf{V}_N(\boldsymbol{\theta})$ *is defined in (2.34),* $\mathbf{U}(\boldsymbol{\theta})$ *is defined in Theorem 3,*

$$\boldsymbol{\Delta} = blockdiag(\frac{1}{r^{(1)}}\boldsymbol{\Delta}^{(1)}, \cdots, \frac{1}{r^{(I)}}\boldsymbol{\Delta}^{(I)}),$$

$$\boldsymbol{\Delta}^{(i)} = blockdiag(\boldsymbol{\Delta}_1^{(i)}, \cdots, \boldsymbol{\Delta}_{L^{(i)}}^{(i)}), \quad i = 1, 2, \cdots, I,$$

$$\boldsymbol{\Delta}_\ell^{(i)} = Var\{v(\mathbf{f}_{\ell j}^{(i)}\mathbf{f}_{\ell j}^{(i)\prime})\} - 2\mathbf{D}_{q_\ell^{(i)}}^{+}(\boldsymbol{\Phi}_\ell^{(i)0} \otimes \boldsymbol{\Phi}_\ell^{(i)0})\mathbf{D}_{q_\ell^{(i)}}^{+\prime}, \quad \ell = 1, 2, \cdots, L^{(i)},$$

$$\mathbf{G} = \begin{pmatrix} \mathbf{G}_1 \\ \mathbf{G}_2 \\ \mathbf{G}_3 \end{pmatrix} = (\mathbf{F}_\infty'\boldsymbol{\Omega}_\infty^{-1}\mathbf{F})_\infty^{-1}\mathbf{F}_\infty'\boldsymbol{\Omega}_\infty^{-1}\boldsymbol{\Lambda},$$

the dimensions of $\mathbf{G}_1$, $\mathbf{G}_2$, *and* $\mathbf{G}_3$ *are* $d_{\boldsymbol{\tau}} \times d_{\boldsymbol{\rho}}$, $d_{\boldsymbol{\rho}} \times d_{\boldsymbol{\rho}}$, *and* $d_{\boldsymbol{\varphi}} \times d_{\boldsymbol{\rho}}$, *respectively,*

$$\boldsymbol{\Lambda} = blockdiag(\frac{1}{r^{(1)}}\boldsymbol{\Lambda}^{(1)}, \cdots, \frac{1}{r^{(I)}}\boldsymbol{\Lambda}^{(I)}), \quad \boldsymbol{\Lambda}^{(i)} = [\boldsymbol{\Lambda}_1^{(i)}, \cdots, \boldsymbol{\Lambda}_{L^{(i)}}^{(i)}],$$

and

$$\Lambda_\ell^{(i)} = \begin{pmatrix} \mathbf{B}_\ell^{(i)}(\boldsymbol{\tau}_0)\boldsymbol{\Gamma}_\ell^{(i)} \\ \mathbf{0} \end{pmatrix}, \quad \boldsymbol{\Gamma}_\ell^{(i)} = E\{\mathbf{f}_{\ell j}^{(i)} v(\mathbf{f}_{\ell j}^{(i)}\mathbf{f}_{\ell j}^{(i)\prime})\}.$$

Proof: Note that

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_{\mathbf{n}}) = \sqrt{n}(\hat{\boldsymbol{\theta}} - \bar{\boldsymbol{\theta}}_{\mathbf{n}}) + \sqrt{n}(\bar{\boldsymbol{\theta}}_{\mathbf{n}} - \boldsymbol{\theta}_{\mathbf{n}}),$$

$$\bar{\boldsymbol{\theta}}_{\mathbf{n}} - \boldsymbol{\theta}_{\mathbf{n}} = (\mathbf{0},\ \boldsymbol{\rho}_{\mathbf{n}}' - \boldsymbol{\rho}_0',\ \mathbf{0})', \tag{2.35}$$

$$\boldsymbol{\rho}_{\mathbf{n}} - \boldsymbol{\rho}_0 = \begin{pmatrix} \boldsymbol{\rho}_{\mathbf{n}}^{(1)} - \boldsymbol{\rho}_0^{(1)} \\ \vdots \\ \boldsymbol{\rho}_{\mathbf{n}}^{(I)} - \boldsymbol{\rho}_0^{(I)} \end{pmatrix}, \quad \boldsymbol{\rho}_{\mathbf{n}}^{(i)} - \boldsymbol{\rho}_0^{(i)} = \begin{pmatrix} \mathrm{v}[\boldsymbol{\Phi}_1^{(i)}(n^{(i)}) - \boldsymbol{\Phi}_1^{(i)0}] \\ \vdots \\ \mathrm{v}[\boldsymbol{\Phi}_{L^{(i)}}^{(i)}(n^{(i)}) - \boldsymbol{\Phi}_{L^{(i)}}^{(i)0}] \end{pmatrix},$$

and that the expansion of $\sqrt{n}(\hat{\boldsymbol{\theta}} - \bar{\boldsymbol{\theta}}_{\mathbf{n}})$ is given in Theorem 2. The leading term in $\sqrt{n}(\hat{\boldsymbol{\theta}} - \bar{\boldsymbol{\theta}}_{\mathbf{n}})$ is a linear function of the quantities in (2.26). Thus, with the finite fourth moments of $\mathbf{f}_{\ell j}^{(i)}$, $\ell = 1, 2, \cdots, L^{(i)}$, the limiting normality of $\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_{\mathbf{n}})$ follows from the central limiting theorem. Also, the limiting covariance matrix $\mathbf{V}_G$ of $\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_{\mathbf{n}})$ can be partitioned as

$$\mathbf{V}_G = \mathbf{V}_G^{(11)} + \mathbf{V}_G^{(22)} + \mathbf{V}_G^{(12)} + \mathbf{V}_G^{(21)},$$

where the superscripts 1 and 2 correspond to $\sqrt{n}(\hat{\boldsymbol{\theta}} - \bar{\boldsymbol{\theta}}_{\mathbf{n}})$ and $\sqrt{n}(\bar{\boldsymbol{\theta}}_{\mathbf{n}} - \boldsymbol{\theta}_{\mathbf{n}})$, respectively. If all $\mathbf{f}_{\ell j}^{(i)}$, $\ell = 1, 2, \cdots, L^{(i)}$, were normally distributed, then the limiting covariance matrix of $\sqrt{n}(\hat{\boldsymbol{\theta}} - \bar{\boldsymbol{\theta}}_{\mathbf{n}})$ would be

$$\tilde{\mathbf{V}}_N = \mathbf{V}_N(\boldsymbol{\theta}_\infty) - \begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{U}(\boldsymbol{\theta}_\infty) \end{pmatrix}.$$

Since $\sqrt{n}(\hat{\boldsymbol{\theta}} - \bar{\boldsymbol{\theta}}_{\mathbf{n}})$ and $\sqrt{n}(\bar{\boldsymbol{\theta}}_{\mathbf{n}} - \boldsymbol{\theta}_{\mathbf{n}})$ are independent under the assumption of normal $\mathbf{f}_{\ell j}^{(i)}$, $\ell = 1, 2, \cdots, L^{(i)} + 1$, we have

$$\tilde{\mathbf{V}}_N = \tilde{\mathbf{V}}_N^{(11)} + \tilde{\mathbf{V}}_N^{(22)}.$$

By Theorem 2 and (2.26), the limiting distribution of $\sqrt{n}(\hat{\boldsymbol{\theta}} - \bar{\boldsymbol{\theta}}_{\mathbf{n}})$ is common for any distribution of $\mathbf{f}_{\ell j}^{(i)}$, $\ell = 1, 2, \cdots, L^{(i)}$, under Assumption 1, and thus, $\tilde{\mathbf{V}}_N^{(11)} = \mathbf{V}_G^{(11)}$. Hence

$$\mathbf{V}_G = \tilde{\mathbf{V}}_N + \mathbf{V}_G - \tilde{\mathbf{V}}_N = \tilde{\mathbf{V}}_N + \mathbf{V}_G^{(22)} - \tilde{\mathbf{V}}_N^{(22)} + \mathbf{V}_G^{(12)} + \mathbf{V}_G^{(12)\prime}.$$

By considering the limiting distributions of $\boldsymbol{\Phi}_\ell^{(i)}(n^{(i)})$ under Assumption 1 and the normal $\mathbf{f}_{\ell j}^{(i)}$, $\ell = 1, 2, \cdots, L^{(i)}$,

$$\mathbf{V}_G^{(22)} - \tilde{\mathbf{V}}_N^{(22)} = \begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Delta} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix}.$$

By (2.24), (2.26), (2.35), and by noting that the limiting covariance between $\mathbf{c}^{(i)} - \boldsymbol{\gamma}^{(i)}(\bar{\boldsymbol{\theta}}_{\mathbf{n}})$ and $\mathrm{v}[\boldsymbol{\Phi}_\ell^{(i)}(n^{(i)})]$ is $r^{(i)-1}\boldsymbol{\Lambda}_\ell^{(i)}$, $\ell = 1, 2, \cdots, L^{(i)}$, we obtain

$$\mathbf{V}_G^{(12)} = (\mathbf{F}_\infty' \boldsymbol{\Omega}_\infty^{-1} \mathbf{F}_\infty)^{-1} \mathbf{F}_\infty' \boldsymbol{\Omega}_\infty^{-1} (\mathbf{0} \quad \boldsymbol{\Lambda} \quad \mathbf{0}) = (\mathbf{0} \quad \mathbf{G} \quad \mathbf{0}) = \begin{pmatrix} \mathbf{0} & \mathbf{G}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{G}_2 & \mathbf{0} \\ \mathbf{0} & \mathbf{G}_3 & \mathbf{0} \end{pmatrix}.$$

Note that the third moment matrix $\boldsymbol{\Gamma}_\ell^{(i)}$ of $\mathbf{f}_{\ell j}^{(i)}$ in $\boldsymbol{\Lambda}_\ell^{(i)}$, $\ell = 1, 2, \cdots, L^{(i)}$; $i = 1, 2, \cdots, I$, comes from the correlation of $\bar{\mathbf{f}}_\ell^{(i)}$ and $\mathrm{v}[\boldsymbol{\Phi}_\ell^{(i)}(n^{(i)})]$, and that the $\mathbf{0}$ matrix in $\boldsymbol{\Lambda}_\ell^{(i)}$ comes from the fact that $\mathrm{v}[\mathbf{S}^{(i)} - \boldsymbol{\Sigma}^{(i)}(\bar{\boldsymbol{\theta}}_{\mathbf{n}})]$ and $\mathrm{v}[\boldsymbol{\Phi}_\ell^{(i)}(n^{(i)})]$ are uncorrelated in the limit. □

The form of $\mathbf{V}_G$ in Theorem 4 shows the difference between $\mathbf{V}_G$ and $\mathbf{V}_N$. The matrix $\boldsymbol{\Delta}$ consists of the fourth-order cumulants of $\mathbf{f}_{\ell j}^{(i)}$, $\boldsymbol{\Delta}_\ell^{(i)}$, and the matrices $\mathbf{G}_1$,

$\mathbf{G}_2$, and $\mathbf{G}_3$ contain the third-order cumulants of $\mathbf{f}_{\ell j}^{(i)}$, $\mathbf{\Gamma}_{\ell}^{(i)}$, for $\ell = 1, 2, \cdots, L^{(i)}$; $i = 1, 2, \cdots, I$.

Up to this point, I populations (or samples from them) were assumed independent. In practice, I samples or populations may be correlated or dependent. For example, the I samples may be I repeated measures of a single population which is believed to evolve over time. Another possibility is a situation where I populations are spatially correlated. It turns out that the inclusion of the fixed latent vector $\mathbf{f}_{L^{(i)}+1,j}^{(i)}$ in model (2.3) provides a treatment of cases with dependent populations or longitudinal studies. If the primary interest in such studies using model (2.3) is in $\boldsymbol{\tau}$, the parameter for relationships and population-specific normal latent vector, then the correct inferences can be obtained by conditioning on the values of the latent vectors correlated over populations and by treating them fixed. For random $\mathbf{f}_{L^{(i)}+1,j}^{(i)}$ with possible dependency over populations, we assume

Assumption 1 *iv*-a) $\{\mathbf{f}_{L^{(i)}+1,j}^{(i)} : i = 1, 2, \cdots, I;\ j = 1, 2, \cdots, n^{(i)}\}$ *is a set of jointly distributed random vectors with any dependency, is independent of* $\{\mathbf{f}_{\ell j}^{(i)} : i = 1, 2, \cdots, I;\ j = 1, 2, \cdots, n^{(i)};\ \ell = 0, 1, \cdots, L^{(i)}\}$, *and satisfies (2.4) and (2.5) with probability one.*

Theorem 5 *Let model (2.3) and Assumptions 1 i), ii), iii), iv-a), 1 v), 2, 3, and 4 hold. Then, as* $n_m \to \infty$,

$$\sqrt{n}(\hat{\boldsymbol{\tau}} - \boldsymbol{\tau}_0) \xrightarrow{L} \mathrm{N}(\mathbf{0},\ \mathbf{V}_{N\boldsymbol{\tau}\boldsymbol{\tau}}(\boldsymbol{\theta}_\infty)),$$

where $\mathbf{V}_{N\boldsymbol{\tau}\boldsymbol{\tau}}(\boldsymbol{\theta}_\infty)$ *is defined in (2.34). In addition,*

$$\mathbf{V}_{N\boldsymbol{\tau}\boldsymbol{\tau}}(\hat{\boldsymbol{\theta}}) \xrightarrow{p} \mathbf{V}_{N\boldsymbol{\tau}\boldsymbol{\tau}}(\boldsymbol{\theta}_\infty).$$

Proof: Consider conditioning on those values of $\mathbf{f}^{(i)}_{L^{(i)}+1,j}$ satisfying (2.4) and (2.5). Then, under Assumption 1 *iv*-a), the set of such values has probability one, and the conditioning does not affect the distribution of other latent vectors. Then, the results of Theorems 2 and 3 hold conditionally for almost all $\mathbf{f}^{(i)}_{L^{(i)}+1,j}$. But, the limiting covariance matrix $\mathbf{V}_{N\boldsymbol{\tau}\boldsymbol{\tau}}(\boldsymbol{\theta}_{\infty})$ of $\sqrt{n}(\hat{\boldsymbol{\tau}} - \boldsymbol{\tau}_0)$ in Theorem 3 depends on $\mathbf{f}^{(i)}_{L^{(i)}+1,j}$ only through the limiting quantities $\boldsymbol{\varphi}^{(i)\infty}_{L^{(i)}+1}$ and $\boldsymbol{\Phi}^{(i)\infty}_{L^{(i)}+1}$ in (2.4) and (2.5). Hence, by taking the expectation of the conditional distribution function of $\sqrt{n}(\hat{\boldsymbol{\tau}} - \boldsymbol{\tau}_0)$, and by taking the limit under the expectation (Dominated Convergence Theorem), it follows that the unconditional distribution function converges to the same limit. The convergence of the estimated covariance matrix follows from the conditional version of Theorem 1. □

Note that Assumption 1 *iv*-a) is very unrestrictive. Thus, almost any dependency over populations or over individuals within a population expressed in terms of ergodic or stationary latent vectors is allowed. Note also that $\mathbf{f}^{(i)}_{L^{(i)}+1,j}$'s do not have to be identically distributed within populations. Hence, the asymptotic standard errors of $\hat{\boldsymbol{\tau}}$ computed under the assumptions of normal data and independent populations are valid also for non-normal data from dependent populations. This provides very simple and correct inferences for $\boldsymbol{\tau}$, the parameter of interest in model checking and inferences, even when the true underlying structure of a problem is extremely complex.

In Theorem 5, the limiting distribution of $\hat{\boldsymbol{\tau}}$ was addressed. For the case with possibly dependent populations expressed in terms of dependent $\mathbf{f}^{(i)}_{L^{(i)}+1,j}$, the inferences for the means of $\mathbf{f}^{(i)}_{L^{(i)}+1,j}$ may be of interest. For the means to make sense, and to discuss the limiting distribution of the mean estimators $\hat{\boldsymbol{\varphi}}^{(i)}_{L^{(i)}+1}$, we need a stronger assumption than Assumption 1 iv-a):

Assumption 1 iv-b) *$\mathbf{f}^{(i)}_{L^{(i)}+1,j}$'s are random vectors independent of $\{\mathbf{f}^{(i)}_{\ell j} : i = 1,2,\cdots,I;\ j = 1,2,\cdots,n^{(i)};\ \ell = 0,1,\cdots,L^{(i)}\}$, and for each given i $\{\mathbf{f}^{(i)}_{L^{(i)}+1,j} : j = 1,2,\cdots,n^{(i)}\}$ is a set of independently and identically distributed random vectors with mean $\boldsymbol{\varphi}^{(i)0}_{L^{(i)}+1}$ and covariance matrix $\boldsymbol{\Phi}^{(i)0}_{L^{(i)}+1}$.*

In this assumption, $\mathbf{f}^{(i)}_{L^{(i)}+1,j}$'s are independent within a population but can still have general unspecified dependency over populations. Clearly, Assumption 1 iv-b) implies Assumption iv-a), and the result for $\boldsymbol{\tau}$ in Theorem 5 holds under Assumption iv-b). In addition, Assumption iv-b) provides a result for the mean estimator $\hat{\boldsymbol{\varphi}}^{(i)}_{L^{(i)}+1}$.

Theorem 6 *Let model (2.3) and Assumptions 1 i), ii), iii), iv-b), v), 2, 3, and 4 hold. Then, as $n_m \to \infty$,*

$$\sqrt{n}(\hat{\boldsymbol{\varphi}}^{(i)}_{L^{(i)}+1} - \boldsymbol{\varphi}^{(i)0}_{L^{(i)}+1}) \xrightarrow{L} \mathrm{N}(\mathbf{0},\ \mathbf{V}_{N\boldsymbol{\varphi}^{(i)}_{L^{(i)}+1}\boldsymbol{\varphi}^{(i)}_{L^{(i)}+1}}(\boldsymbol{\theta}_0)), \quad i = 1,2,\cdots,I,$$

where $\mathbf{V}_{N\boldsymbol{\varphi}^{(i)}_{L^{(i)}+1}\boldsymbol{\varphi}^{(i)}_{L^{(i)}+1}}(\boldsymbol{\theta}_0)$ is the part of $\mathbf{V}_N(\boldsymbol{\theta}_0)$ in (2.32) corresponding to $\boldsymbol{\varphi}^{(i)}_{L^{(i)}+1}$, and $\boldsymbol{\theta}_0$ is as defined in (2.31) with $\boldsymbol{\varphi}^{(i)0}_{L^{(i)}+1}$ and covariance matrix $\boldsymbol{\Phi}^{(i)0}_{L^{(i)}+1}$ of Assumption 1 iv-b). In addition,

$$\mathbf{V}_{N\boldsymbol{\varphi}^{(i)}_{L^{(i)}+1}\boldsymbol{\varphi}^{(i)}_{L^{(i)}+1}}(\hat{\boldsymbol{\theta}}) \xrightarrow{p} \mathbf{V}_{N\boldsymbol{\varphi}^{(i)}_{L^{(i)}+1}\boldsymbol{\varphi}^{(i)}_{L^{(i)}+1}}(\boldsymbol{\theta}_0).$$

Proof: Following the steps of the proof of Theorem 3 with $\boldsymbol{\theta}_\infty$ replaced by $\boldsymbol{\theta}_0$, and making appropriate adjustments for random $\mathbf{f}^{(i)}_{L^{(i)}+1,j}$ satisfying Assumption 1 iv-b), we obtain

$$\sqrt{n}(\hat{\boldsymbol{\varphi}}^{(i)}_{L^{(i)}+1} - \bar{\mathbf{f}}^{(i)}_{L^{(i)}+1}) \xrightarrow{L} \mathrm{N}(\mathbf{0},\ \mathbf{V}_{N\boldsymbol{\varphi}^{(i)}_{L^{(i)}+1}\boldsymbol{\varphi}^{(i)}_{L^{(i)}+1}}(\boldsymbol{\theta}_0) - \frac{1}{r^{(i)}}\boldsymbol{\Phi}^{(i)0}_{L^{(i)}+1}), \quad i = 1,2,\cdots,I.$$

Under Assumption 1 iv-b),

$$\sqrt{n}(\bar{\mathbf{f}}^{(i)}_{L^{(i)}+1} - \boldsymbol{\varphi}^{(i)0}_{L^{(i)}+1}) \xrightarrow{L} \mathbf{N}(\mathbf{0}, \frac{1}{r^{(i)}}\boldsymbol{\Phi}^{(i)0}_{L^{(i)}+1}), \quad i = 1, 2, \cdots, I.$$

The first result follows from the fact that, by the form of (2.26), $\sqrt{n}(\hat{\boldsymbol{\varphi}}^{(i)}_{L^{(i)}+1} - \bar{\mathbf{f}}^{(i)}_{L^{(i)}+1})$ and $\sqrt{n}(\bar{\mathbf{f}}^{(i)}_{L^{(i)}+1} - \boldsymbol{\varphi}^{(i)0}_{L^{(i)}+1})$ are independent in the joint limiting distribution. □

According to Theorem 6, if the random latent vectors $\mathbf{f}^{(i)}_{L^{(i)}+1,j}$ with general dependency over i form a random sample for each i, then the inferences for the mean of $\mathbf{f}^{(i)}_{L^{(i)}+1,j}$ for each i can be made correctly using the corresponding parts of the limiting covariance matrix $\mathbf{V}_N(\boldsymbol{\theta}_0)$ for the normal-independence model. Theorems 5 and 6 state the results for $\hat{\boldsymbol{\tau}}$ and $\hat{\boldsymbol{\varphi}}^{(i)}_{L^{(i)}+1}$ separately. It can be shown that under Assumption 1 iv-b) the joint limiting distribution of $\hat{\boldsymbol{\tau}}$ and $\hat{\boldsymbol{\varphi}}^{(i)}_{L^{(i)}+1}$ for each i is the same as that under the normal-independence model, although the limiting covariance between $\hat{\boldsymbol{\tau}}$ and $\hat{\boldsymbol{\varphi}}^{(i)}_{L^{(i)}+1}$ is not of practical interest.

2.5 Simulation Study

In this section, the results from a simulation study are presented for a moderately complicated latent variable model. For similar studies using much simpler models, see Satorra (1993b) and Papadopoulos and Amemiya (1994). We consider a two-population recursive system of simultaneous equations with errors in the explanatory variables

$$
\begin{array}{rlllll}
y_{1j}^{(1)} = & \alpha_1 + & & & \gamma_1 \xi_j^{(1)} + & \zeta_{1j}^{(1)} \\
y_{2j}^{(1)} = & \alpha_2 + & \beta_{21} y_{1j}^{(1)} + & & \gamma_2 \xi_j^{(1)} + & \zeta_{2j}^{(1)} \\
y_{3j}^{(1)} = & \alpha_3^{(1)} + & & \beta_{32}^{(1)} y_{2j}^{(1)} + & & \zeta_{3j}^{(1)} \\
x_j^{(1)} = & & & & \xi_j^{(1)} + & \delta_j^{(1)}
\end{array}
$$

$$
\begin{array}{rllll}
y_{1j}^{(2)} = & \alpha_1 + & & \gamma_1 \xi_j^{(2)} + & \zeta_{1j}^{(2)} \\
y_{2j}^{(2)} = & \alpha_2 + & \beta_{21} y_{1j}^{(2)} + & \gamma_2 \xi_j^{(2)} + & \zeta_{2j}^{(2)} \\
x_j^{(2)} = & & & \xi_j^{(2)} + & \delta_j^{(2)}
\end{array}
$$

This model is also a special case of the LISREL model with $\mathbf{y}^{(i)} = \boldsymbol{\eta}^{(i)}$, and can be written in the form of model (2.3) as indicated in Section 2.2. Note that the number of observed variables is different for the two populations. Four measurements $y_{1j}^{(1)}$, $y_{2j}^{(1)}$, $y_{3j}^{(1)}$, and $x_j^{(1)}$ are taken from the first population ($p^{(1)} = 4$) and three measurements $y_{1j}^{(2)}$, $y_{1j}^{(2)}$, $x_j^{(2)}$ from the second ($p^{(2)} = 3$). The parameters α_1, α_2, β_{21}, γ_1, and γ_2 do not depend on i. That is, α_1, α_2, β_{21}, γ_1, and γ_2 are common for the two populations. The sample sizes are 550 and 500 for the first and the second populations, respectively ($n^{(1)} = 550$, $n^{(2)} = 500$). Such a set of two samples was generated 1,000 times.

Two cases are considered. The unobservable variables $\xi_j^{(i)}$, $i = 1, 2$, were treated as fixed in Case 1 and as non-normal in Case 2. In both cases, $\xi_j^{(1)}$ and $\xi_j^{(2)}$ are related (correlated over populations), and were generated by adjusted chi-square random variables with one degree of freedom. In Case 1, one sample of ($\xi_j^{(1)}$, $\xi_j^{(2)}$) was generated with sample means, variances, and covariance

$$\bar{\xi}^{(1)} = \frac{1}{550} \sum_{j=1}^{550} \xi_j^{(1)} = 4.995, \quad \bar{\xi}^{(2)} = 9.913,$$

$$s_{\xi^{(1)}\xi^{(1)}} = \frac{1}{549}\sum_{j=1}^{550}(\xi_j^{(1)} - \bar{\xi}^{(1)})^2 = 1.814, \quad s_{\xi^{(2)}\xi^{(2)}} = 1.699,$$

$$s_{\xi^{(1)}\xi^{(2)}} = \frac{1}{499}\sum_{j=1}^{500}(\xi_j^{(1)} - \bar{\xi}^{(1)})(\xi_j^{(2)} - \bar{\xi}^{(2)}) = 1.319,$$

and this one set of $(\xi_j^{(1)},\ \xi_j^{(2)})$ was used in all 1,000 Monte Carlo samples. In Case 2, 1,000 independent samples were generated for $\{\xi_j^{(1)},\ j = 1,2,\cdots,550;\ \xi_j^{(2)},\ j = 1,2,\cdots,500\}$. The true means, variances, and covariance of $\xi_j^{(1)}$ and $\xi_j^{(2)}$ are $\mu^0_{\xi^{(1)}} = 5$, $\mu^0_{\xi^{(2)}} = 10$, $\sigma^0_{\xi^{(1)}\xi^{(1)}} = 2$, $\sigma^0_{\xi^{(2)}\xi^{(2)}} = 2$, and $\sigma^0_{\xi^{(1)}\xi^{(2)}} = 1.4$.

In both cases, 1,000 samples were generated for independent $\zeta_{1j}^{(1)}$, $\zeta_{2j}^{(1)}$, $\zeta_{3j}^{(1)}$, $\delta_j^{(1)}$, $\zeta_{1j}^{(2)}$, $\zeta_{2j}^{(2)}$, and $\delta_j^{(2)}$, $j = 1,2,\cdots,n^{(i)}$. The errors $\delta_j^{(1)}$ and $\delta_j^{(2)}$ are normally distributed with known mean 0 and common unknown variance $\sigma_{\delta\delta}$. The variables $\zeta_{1j}^{(1)}$, $\zeta_{2j}^{(1)}$, $\zeta_{3j}^{(1)}$, $\zeta_{1j}^{(2)}$ and $\zeta_{2j}^{(2)}$ are independent adjusted chi-square random variables with one degree of freedom with known means 0 and unknown variances $\mathrm{Var}\{\zeta_{kj}^{(i)}\} = \sigma_{\zeta_k^{(i)}\zeta_k^{(i)}}$. In accordance with the notation of this paper,

$$\boldsymbol{\theta} = (\boldsymbol{\tau}', \boldsymbol{\rho}', \boldsymbol{\varphi}')',$$

$$\boldsymbol{\tau} = (\alpha_1,\ \alpha_2,\ \alpha_3^{(1)},\ \beta_{21},\ \beta_{32}^{(1)},\ \gamma_1,\ \gamma_2,\ \sigma_{\delta\delta})',$$

$$\boldsymbol{\rho} = (\sigma_{\zeta_1^{(1)}\zeta_1^{(1)}},\ \sigma_{\zeta_2^{(1)}\zeta_2^{(1)}},\ \sigma_{\zeta_3^{(1)}\zeta_3^{(1)}},\ \sigma_{\zeta_1^{(2)}\zeta_1^{(2)}},\ \sigma_{\zeta_2^{(2)}\zeta_2^{(2)}})'$$

in both cases, and

$$\boldsymbol{\varphi} = (\bar{\xi}^{(1)},\ s_{\xi^{(1)}\xi^{(1)}},\ \bar{\xi}^{(2)},\ s_{\xi^{(2)}\xi^{(2)}})'$$

in Case 1 and

$$\boldsymbol{\varphi} = (\mu_{\xi^{(1)}},\ \sigma_{\xi^{(1)}\xi^{(1)}},\ \mu_{\xi^{(2)}},\ \sigma_{\xi^{(2)}\xi^{(2)}})'$$

in Case 2. The true values of $\boldsymbol{\tau}$ and $\boldsymbol{\rho}$ are given in Table 2.1. Note that in Case 1 the $\xi_j^{(i)}$, $i = 1,2$ are treated as fixed satisfying Assumption 1 *iv*) while in Case 2 the

$\xi_j^{(i)}$, $i = 1, 2$ are treated as random satisfying Assumption 1 *iv*-b). Thus, Theorems 3 and 4 apply for Case 1, and Theorems 5 and 6 for Case 2.

Figure 2.1 presents histograms and non-parametric density estimates for some of the observed variables for one of 1,000 generated samples for Case 2. The plots indicates that the distributions of the observations are skewed with heavy tails and deviate considerably from normality. Such observations appear often in practice.

Tables 2.1 and 2.2 summarize the results of the simulation studies for Cases 1 and 2, respectively. In both tables, the Monte Carlo means (MC-mean) and standard deviations (MC-s.e.) of the parameter estimates are given, and two types of asymptotic standard error formulas are presented. The asymptotic standard errors are evaluated at the limiting true values in Case 1 and at the true values in Case 2. The general model standard errors (G-a.s.e.) are computed based on an appropriate matrix formula that is asymptotically correct for all the parameters. For Case 1 in Table 2.1 with fixed $\xi_j^{(i)}$ G-a.s.e. was computed based on $n^{-1}\mathbf{V}_G(\boldsymbol{\theta}_\infty)$, where $\mathbf{V}_G(\boldsymbol{\theta})$ is defined in Theorem 4, $nr^{(1)} = 550$, $nr^{(2)} = 500$, and $\boldsymbol{\theta}_\infty = (\boldsymbol{\tau}_0', \ \boldsymbol{\rho}_0', \ 5, \ 2, \ 10, \ 2)'$. For Case 2 in Table 2.2 with random $\xi_j^{(i)}$, the matrix $n^{-1}\mathbf{V}_G(\boldsymbol{\theta}_0)$ was used to compute asymptotically correct standard errors for the estimates of $\boldsymbol{\tau}$ and $\boldsymbol{\rho}$. The G-a.s.e.'s for $\boldsymbol{\varphi}$ in Case 2 were computed by adding the fourth-order cumulants of $\xi_j^{(1)}$ and $\xi_j^{(2)}$ divided by $n^{(i)}$ to the asymptotic variances of $\hat{\sigma}_{\xi^{(1)}\xi^{(1)}}$ and $\hat{\sigma}_{\xi^{(2)}\xi^{(2)}}$ in $n^{-1}\mathbf{V}_N(\boldsymbol{\theta}_0)$ of (2.32). For both cases, the knowledge of the third and fourth order moments of the unobservable non-normal variables was used in computing G-a.s.e.'s. Thus, G-a.s.e.'s are the correct asymptotic standard errors for this experiment, but can not be obtained in practice without information on the distributional form of latent variables. The last column of each table gives the asymptotic standard errors (N-a.s.e.) under the normal-independence model computed based on $n^{-1}\mathbf{V}_N(\boldsymbol{\theta}_\infty)$ for Case 1 and on $n^{-1}\mathbf{V}_N(\boldsymbol{\theta}_0)$ for Case 2. The theoretical results in the previous section showed that the G-a.s.e. and N-a.s.e. are common for some parameters. For Case 1 in Table 2.1, Theorem 3 showed that the G-a.s.e. and N-a.s.e. are the same for $\boldsymbol{\tau}$-parameters, and that the N-a.s.e. is larger than the G-a.s.e. for $\boldsymbol{\varphi}$-parameters by the amount corresponding to the $\mathbf{U}(\boldsymbol{\theta}_\infty)$ term. For Case 2 in Table 2.2 with correlated $\xi_j^{(1)}$ and $\xi_j^{(2)}$, Theorems 5 and 6 proved that the N-a.s.e. is equal to the G-a.s.e. for $\boldsymbol{\tau}$-parameters and $\mu_{\xi^{(i)}}$, $i = 1, 2$. In both tables, the bold-faced entries in the N-a.s.e. column correspond to those without theoretical equivalence to the G-a.s.e.. Note in Tables 2.1 and 2.2 that the G-a.s.e. is an excellent approximation to the simulated MC-s.e., and thus the normal-independence model N-a.s.e. is practically useful when the theoretical asymptotic correctness holds. The tables also show that the normal-independence model maximum likelihood estimators applied to non-normal and dependent models have very small bias for all parameters.

For more detailed assessment of the bias of the normal-independence estimator in terms of sample sizes, various values of $n^{(1)}$ and $n^{(2)}$ were used for Case 1 with the above parameterization. For three cases $(n^{(1)}, n^{(2)}) = (110, 100), (200, 150), (300, 50)$, Table 2.3 reports the percentage relative bias for a number of parameters in $\boldsymbol{\tau}$. The percentage relative bias is

$$100(\text{MC mean} - \text{true value})/\text{true value}.$$

The last row of Table 2.3 gives the average of the absolute values of the percentage relative biases for the eight parameters in $\boldsymbol{\tau}$. Overall, the bias of the normal-independence maximum likelihood estimator is small, and is negligible even with $n^{(i)}$ as low as 150-200. Note that the bias is very small even when one of the sample sizes is as small as 50. For larger $n^{(1)}$ and $n^{(2)}$ the percentage relative bias becomes even smaller.

To see how well the asymptotic standard errors approximate the true standard deviations, the model for Case 1 was repeated for different combinations of $n^{(1)}$ and $n^{(2)}$. Table 2.4 gives a summary of the results in terms of the percentage relative difference, i.e.,

$$100(\text{asymptotic s.e.} - \text{MC s.e.})/\text{MC s.e.}.$$

The last row of the table presents the average of the absolute percentage relative differences for the eight parameters in $\boldsymbol{\tau}$. For the $\boldsymbol{\tau}$-parameters, the asymptotic s.e. can be computed using the normal-independence case formula (N-a.s.e.). The relative difference in standard errors is less than 10% for $n^{(1)}$ and $n^{(2)}$ as small as 150-200, and becomes very small when $n^{(1)}$ and $n^{(2)}$ are larger than 250. The last column of Table 2.4 corresponding to a case with very uneven $n^{(1)}$ and $n^{(2)}$ shows that the asymptotic approximation is reasonable even when one of the sample sizes is relatively small. Tables similar to Tables 2.3 and 2.4 were also obtained for Case 2 with random $\xi^{(i)}$. The overall results for Case 2 were nearly identical to those for Case 1.

In summary, model (2.3) with the errors-in-variables parameterization can formulate the multi-population analysis in a meaningful fashion. The corresponding statistical analysis under the pseudo normal-independence model gives a simple and correct way to conduct statistical inferences about the parameter vector $\boldsymbol{\tau}$ without specifying a distributional form or dependency structure over populations. In practice, $\boldsymbol{\tau}$ often contains all the parameters of direct interest. The asymptotic covariance matrix and standard errors can be readily computed using the existing packages, and provide a good approximation in moderately sized samples.

Table 2.1: Simulation Result for Case 1. Fixed $\xi_j^{(i)}$ $i = 1, 2$, $n^{(1)} = 550$, $n^{(2)} = 500$, and 1000 replications. Two types of asymptotic standard errors, G-a.s.e. under the general model and N-a.s.e. under the normal-independence model, are to be compared to the Monte Carlo standard deviations (MC-s.e.). The boldface numbers indicate the lack of theoretical correctness of the standard errors.

	$\boldsymbol{\theta}$	true value	MC-mean	MC-s.e.	G-a.s.e.	N-a.s.e.
	a_1	1.000	1.0022	0.086	0.084	0.084
	a_2	2.000	1.9921	0.142	0.142	0.142
	$a_3^{(1)}$	-1.000	-0.9966	0.140	0.135	0.135
$\boldsymbol{\tau}$	β_{21}	-1.000	-0.9951	0.082	0.085	0.085
	$\beta_{32}^{(1)}$	1.000	0.9997	0.012	0.012	0.012
	γ_1	-1.000	-1.0004	0.011	0.011	0.011
	γ_2	1.000	1.0053	0.087	0.090	0.090
	$\sigma_{\delta\delta}$	0.300	0.2994	0.039	0.041	0.041
	$\sigma_{\zeta_1^{(1)}\zeta_1^{(1)}}$	0.500	0.4986	0.086	0.088	**0.055**
	$\sigma_{\zeta_2^{(1)}\zeta_2^{(1)}}$	0.600	0.6001	0.110	0.113	**0.060**
$\boldsymbol{\rho}$	$\sigma_{\zeta_3^{(1)}\zeta_3^{(1)}}$	0.700	0.6991	0.111	0.113	**0.042**
	$\sigma_{\zeta_1^{(2)}\zeta_1^{(2)}}$	0.500	0.4982	0.101	0.102	**0.057**
	$\sigma_{\zeta_2^{(2)}\zeta_2^{(2)}}$	0.600	0.5982	0.106	0.107	**0.062**
	$\bar{\xi}^{(1)}$	4.995	4.9950	0.024	0.023	**0.065**
$\boldsymbol{\varphi}$	$s_{\xi^{(1)}\xi^{(1)}}$	1.814	1.8423	0.050	0.051	**0.131**
	$\bar{\xi}^{(2)}$	9.913	9.9123	0.025	0.024	**0.068**
	$s_{\xi^{(2)}\xi^{(2)}}$	1.699	1.6980	0.050	0.053	**0.137**

Table 2.2: Simulation Result for Case 2. Correlated $\xi_j^{(i)}$ $i = 1, 2$, $n^{(1)} = 550$, $n^{(2)} = 500$, and 1000 replications. Two types of asymptotic standard errors, G-a.s.e. under the general model and N-a.s.e. under the normal-independence model, are to be compared to the Monte Carlo standard deviations (MC-s.e.). The boldface numbers indicate the lack of theoretical correctness of the standard errors.

	θ	true value	MC-mean	MC-s.e.	G-a.s.e.	N-a.s.e.
	a_1	1.000	1.0011	0.086	0.084	0.084
	a_2	2.000	1.9957	0.143	0.142	0.142
	$a_3^{(1)}$	-1.000	-1.0001	0.137	0.135	0.135
τ	β_{21}	-1.000	-0.9981	0.084	0.085	0.085
	$\beta_{32}^{(1)}$	1.000	1.0000	0.012	0.012	0.012
	γ_1	-1.000	-1.0002	0.011	0.011	0.011
	γ_2	1.000	1.0024	0.089	0.090	0.090
	$\sigma_{\delta\delta}$	0.300	0.2979	0.040	0.041	0.041
	$\sigma_{\zeta_1^{(1)}\zeta_1^{(1)}}$	0.500	0.5007	0.084	0.088	**0.055**
	$\sigma_{\zeta_2^{(1)}\zeta_2^{(1)}}$	0.600	0.5995	0.110	0.113	**0.060**
ρ	$\sigma_{\zeta_3^{(1)}\zeta_3^{(1)}}$	0.700	0.7003	0.115	0.113	**0.042**
	$\sigma_{\zeta_1^{(2)}\zeta_1^{(2)}}$	0.500	0.5018	0.102	0.102	**0.057**
	$\sigma_{\zeta_2^{(2)}\zeta_2^{(2)}}$	0.600	0.6016	0.104	0.107	**0.062**
	$\mu_{\xi^{(1)}}$	5.000	4.9965	0.064	0.065	0.065
φ	$\sigma_{\xi^{(1)}\xi^{(1)}}$	2.000	1.9862	0.313	0.326	**0.131**
	$\mu_{\xi^{(2)}}$	10.000	9.9967	0.069	0.068	0.068
	$\sigma_{\xi^{(2)}\xi^{(2)}}$	2.000	1.9933	0.259	0.342	**0.137**

Table 2.3: Percentage relative bias. 100×bias/true value. The last row gives the average of the absolute percentage relative biases for the eight parameters in $\boldsymbol{\tau}$.

$n^{(1)}$ $n^{(2)}$	110 100	200 150	300 50
a_1	0.71	0.10	0.81
$a_3^{(1)}$	-0.07	-0.49	-0.63
β_{21}	-4.62	-2.24	-2.88
$\beta_{32}^{(1)}$	-0.04	-0.06	-0.03
γ_2	4.97	2.32	3.09
Average for $\boldsymbol{\tau}$	1.93	0.87	1.28

Table 2.4: Percentage relative difference between Monte Carlo and asymptotic standard errors. 100×(a.s.e.−MC s.e.)/MC s.e.. The last row gives the average of the absolute percentage relative differences for the eight parameters in $\boldsymbol{\tau}$.

$n^{(1)}$ $n^{(2)}$	110 100	200 150	300 250	550 500	300 50
a_1	-5.0	7.2	3.4	-3.0	-0.9
$a_3^{(1)}$	-0.2	-8.0	-1.2	-3.6	-0.6
β_{21}	-24.0	-7.8	2.0	3.7	-12.5
$\beta_{32}^{(1)}$	3.8	-9.2	-2.7	-3.8	-1.6
γ_2	-24.0	-7.1	2.7	3.2	-13.2
Average for $\boldsymbol{\tau}$	10.8	7.2	3.2	3.0	6.0

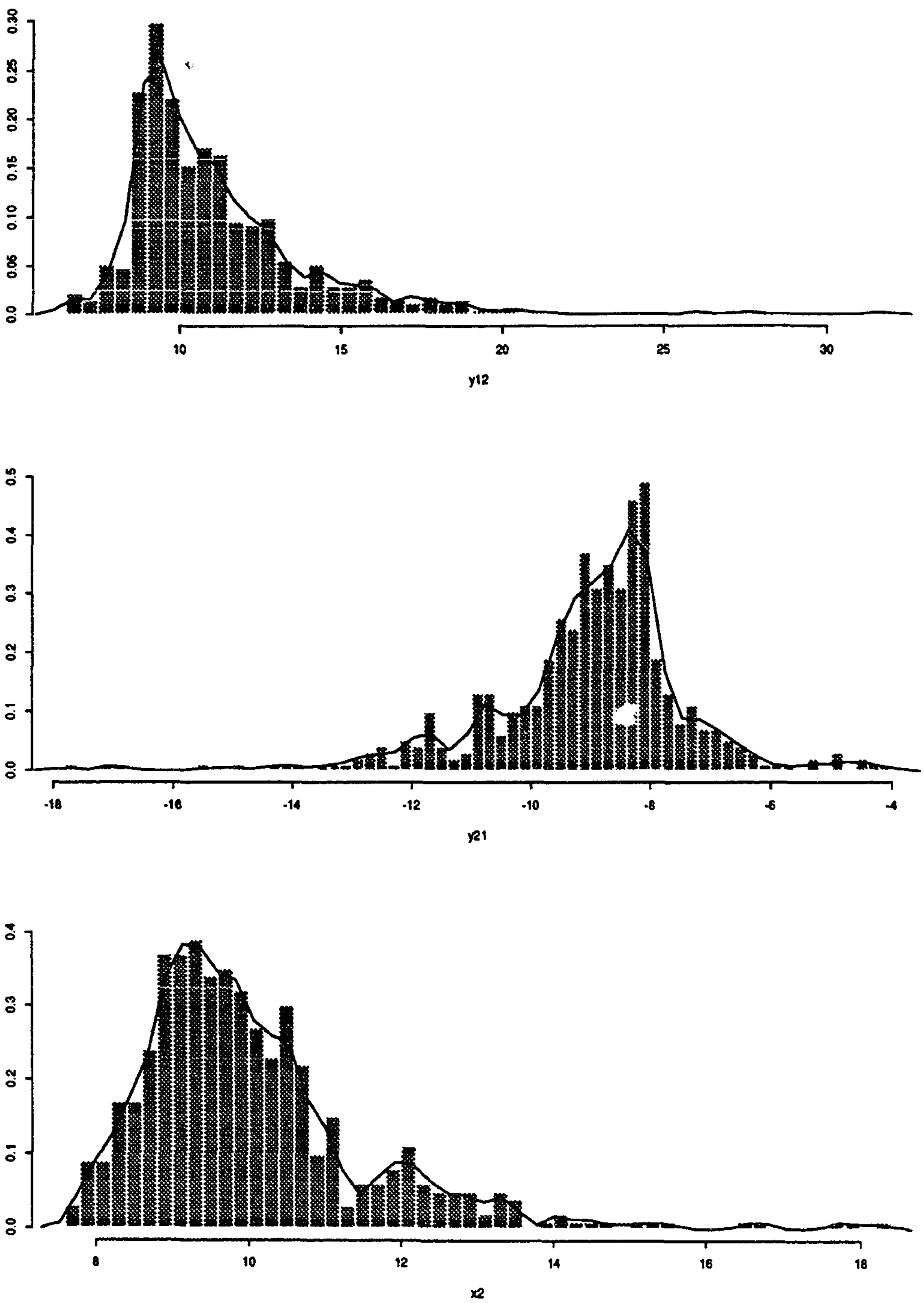


Figure 2.1: Histograms and Non-parametric Density Estimates for the Observed Variables $(y_2^{(1)}, y_1^{(2)}, x^{(2)})$.

2.6 References

Amemiya, Y (1985). On the goodness-of-fit tests for linear statistical relationships. *Technical Report No. 10, Econometric Workshop, Stanford University* Stanford, California.

Amemiya, Y (1986). Multivariate functional and structural relationships with general error covariance structure. *Preprint Series No. 86-3, Department of Statistics, Iowa State University,* Ames, Iowa.

Amemiya, Y., Fuller W.A. and Pantula, S.G. (1987). The asymptotic distributions of some estimators for a factor analysis model. *Journal of Multivariate Analysis*, **22**, 51-64.

Amemiya, Y. and Anderson, T.W. (1990). Asymptotic chi-square tests for a large class of factor analysis models. *The Annals of Statistics*, **18**, 1453-1463.

Anderson, T.W. (1987). Multivariate linear relations. *Proceedings of the Second International Conference in Statistics,* edited by Pukkila T. and Puntanen S., University of Tampere, Finland. 9-36.

Anderson, T.W. (1989). Linear latent variable models and covariance structures. *Journal of Econometrics*, **41**, 91-119.

Anderson, T.W. and Amemiya, Y.(1988). The asymptotic normal distribution of estimators in factor analysis under general conditions. *The Annals of Statistics*, **16**, 759-771.

Bentler, P.M. (1983). Some contributions to efficient statistics in structural models: Specification and estimation of moment structures. Psychometrika, **48**, 493-517.

Bentler, P.M. (1989). *EQS Structural Equations Program Manual.* BMDP Statistical Software, Inc., Los Angeles.

Browne, M.W. (1984). Asymptotically distribution-free methods for the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology*, **37**, 62-83.

Bollen, P.A. (1989). *Structural Equations with Latent Variables.* John Wiley and Sons, New York.

Browne, M.W. (1987). Robustness in statistical inference in factor analysis and related models. *Biometrika*, **74**, 375-384.

Browne, M.W. (1990). Asymptotic robustness of normal theory methods for the analysis of latent curves. *Contemporary Mathematics*, **112**, 211-225.

Browne, M.W. and Shapiro, A. (1988). Robustness of normal theory methods in the analysis of linear latent variate models. *British Journal of Mathematical and Statistical Society*, **41**, 193-208.

Chou, C.P., Bentler, P.M., and Satorra, A. (1991). Scaled test statistics and robust standard errors for non-normal data in covariance structure analysis: A Monte Carlo study. *The British Journal of Mathematical and Statistical Psychology*. **44**, 347-358.

Fuller, W.A. (1987). *Measurement Error Models.* John Wiley and Sons, New York.

Jõreskog, K. (1971). Simultaneous factor analysis in several populations. *Psychometrika*. **47**, 297-308.

Jõreskog, K. and Sõrbom, D. (1989). *LISREL 7; A Guide to the Program and Applications.* 2nd ed., SPSS INC., Chicago.

Kano, Y. (1986). Conditions on consistency of estimators in covariance structure model. *Journal of the Japan Statistical Society*. **16**, 75-80.

Lee, S.Y. and Tsui, K.L. (1982). Covariance structure analysis in several populations. *Psychometrika*, **55**, 107-122.

Magnus, J. and Neudecker, H. (1988). *Matrix Differential Calculus.* John Wiley and Sons, New York.

Mooijaart, A. and Bentler, P.M. (1991). Robustness of normal theory statistics in structural equation models. *Statistica Neerlandica* **45**, 159-171.

Muthén, B. (1989). Multiple group structural modeling with non-normal continuous variables. *British Journal of Mathematical and Statistical Psychology*, **42**, 55-62.

Muthén, B. and Kaplan, D. (1992). A comparison of some methodologies for the factor analysis of non-normal Likert variables: A note on the size of the model. *The British Journal of Mathematical and Statistical Psychology*. **45**, 19-30.

Papadopoulos, S. and Amemiya, Y. (1994). Asymptotic robustness for the structural equation analysis of several populations. *ASA Proceedings*, Business and Economics Statistics Section, 65-70.

Papadopoulos, S. and Amemiya, Y. (1995). On factor analysis of longitudinal data. *ASA Proceedings*, Biometrics Statistics Section.

SAS Institute Inc. (1990). *SAS/STAT User's Guide, Version 6, Fourth Edition, Volume 1,* Cary, NC: SAS Institute Inc.

Satorra, A. (1992). Asymptotic robust inferences in the analysis of mean and covariance structures. *Sociological Methodology,* edited by Marsden, P.V., 249-278.

Satorra, A. (1993a). Multi-sample analysis of moment-structures: Asymptotic validity of inferences based on second-order moments. *Statistical Modeling and Latent Variables,* edited by Haagen, K., Bartholomew, D.J. and Deistler, M.. Amsterdam: Elsevier. Forthcoming.

Satorra, A. (1993b). Asymptotic robust inferences multi-sample analysis of augmented-moment structures. *Multivariate Analysis: Future Directions 2*, edited by Cuadras, C.M. and Rao, C.R., 211-229.

Satorra, A. (1994). On asymptotic robustness in multiple-group analysis of multivariate relations. Paper presented at *Latent Variable Modeling with Applications to Causality*, March 19-20, Los Angeles.

Satorra, A and Bentler, P.M. (1990). Model conditions for asymptotic robustness in the analysis of linear relations. *Computational Statistics and Data Analysis*, **10**, 235-249.

Shapiro, A. (1984). A note on the consistency of estimators in the analysis of moment structures. *British Journal of Mathematical and Statistical Psychology.* **37**, 84-88.

Shapiro, A. (1987). Robustness properties of the MDF analysis of moment structures. *South African Statistical Journal*, **21**, 39-62.

3. ASYMPTOTIC ROBUSTNESS FOR THE STRUCTURAL EQUATION ANALYSIS OF SEVERAL POPULATIONS

A paper published in the 1994 ASA Proceedings
of the Business and Economics Statistics Section [1]

Savas Papadopoulos and Yasuo Amemiya

3.1 Abstract

Structural equation analysis is considered when observations are taken from several populations. For a single population case, it is known that some of the asymptotic standard errors obtained under the assumption of normality are valid for non-normal and fixed latent variables. This asymptotic result is extended to the multi-population case. The assumption of independent samples or populations is relaxed, and a general approach to the multiple correlated samples is introduced. A simulation study is conducted to assess the usefulness of the asymptotic results in finite samples.

[1] *Key words:* Correct Standard errors, fixed and non-normal latent variables, mean and covariance structure, maximum normal likelihood.

3.2 Introduction

In social and behavioral sciences, the structural equation analysis has been very popular. The model for such analysis consists of two parts: The structural equation expresses the relationship of interest between two latent or unobservable vector variables $\boldsymbol{\eta}$, and $\boldsymbol{\xi}$

$$\boldsymbol{\eta} = \boldsymbol{\alpha} + \mathbf{B}\boldsymbol{\eta} + \boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\zeta}, \tag{3.1}$$

where $\boldsymbol{\zeta}$ is the error in the equation with zero mean. The unobservable $\boldsymbol{\eta}$ and $\boldsymbol{\xi}$ are related to the observations through the linear measurement model

$$\mathbf{y} = \boldsymbol{\tau}_y + \boldsymbol{\Lambda}_y\boldsymbol{\eta} + \boldsymbol{\epsilon}, \tag{3.2}$$

$$\mathbf{x} = \boldsymbol{\tau}_x + \boldsymbol{\Lambda}_x\boldsymbol{\xi} + \boldsymbol{\delta}, \tag{3.3}$$

where $\boldsymbol{\epsilon}$ and $\boldsymbol{\delta}$ are measurement errors. The parameters $\boldsymbol{\alpha}$, $\mathbf{B}$, $\boldsymbol{\Gamma}$, $\boldsymbol{\tau}_y$, $\boldsymbol{\tau}_x$ $\boldsymbol{\Lambda}_y$, and $\boldsymbol{\Lambda}_x$ are unknown, but some elements are either known or satisfy equality constrains based on the subject-matter theory or on the identification condition. The model consisting of (3.1)-(3.3) has some widely used important special cases. A model consisting of (3.3) is the factor analytic model, while (3.1) and (3.2) constitute the second order factor analysis. Also, (3.2) and (3.3), when $\boldsymbol{\eta}$ is equal to $\boldsymbol{\xi}$ and $\boldsymbol{\tau}_x$ and $\boldsymbol{\Lambda}_x$ are dropped, constitute the measurement error model. Structural equation models and their special cases are widely used in economics, marketing, psychology, sociology and education. For a general introduction and description, see, Anderson and Rubin [6], Fuller [11], Bollen [8], Bentler [7], and Jöreskog and Sörbom [13].

There are situations where the interest exists in fitting structural equations in several populations and analyzing the corresponding several samples simultaneously.

If the structural equations in the several populations do not have common parameters, then the corresponding samples can be analyzed independently. The case where all the parameters are common over populations is the case of one population. If some parameters are common and some parameters are different over populations, a special technique is required to carry out the analysis of several samples together. Some statistical computer packages such as LISREL can be used for such analysis. But, these packages assume that the populations are independent and normal, and that the samples are simple random samples with no repeated or cluster sampling structure. For a discussion of the structural equation analysis of several populations see, e.g., Jöreskog [12, 13], Lee and Tsui [14], Muthén [16], and Satorra [18, 19].

An area of research which attracted intense interest and made a large impact in structural equation modeling is the so-called asymptotic robustness study. With the wide availability of normality based computer packages, it is important to understand the extent of validity of the statistical inference procedures based on the quantities computed by the packages. It turns out that many of the large sample inference procedures are valid under much weaker assumptions than normality and random sampling. A large amount of important research has been conducted in this area. Amemiya, Fuller and Pantula [1] considered exploratory factor analysis and proved that the asymptotic distribution of the estimated factor loadings and error variances is the same for fixed, non-normal and normal factors under the assumption that the errors are normal. Anderson and Amemiya [5] extended the above results to confirmatory factor analysis and non-normal errors. Browne and Shapiro [10] considered a general linear model with independent normal and non-normal latent variables. Anderson [3] considered the model of Browne and Shapiro with non-stochastic latent variables.

Amemiya and Anderson [2], Anderson [4], and Browne and Shapiro [10] showed the asymptotic robustness of the goodness of fit statistics. Browne [9] and Satorra [17] considered the model of Browne and Shapiro [10] with additional mean structure. All of these papers concentrated on the case of a single population. Satorra [18, 19] extended some of the above results to the multi-population case. In this paper, we consider the asymptotic robustness for a very general multi-population case where the latent vectors may include fixed and non-normal variables, the populations may be related, and the model may contain the mean structure.

In Section 3.3, we introduce the notation, the models used in this paper, and the estimation procedure. Section 3.4 presents the main results without proofs. The proofs and a more detailed discussion are available elsewhere. A simulation study is given in Section 3.5.

3.3 Notation and Model

Suppose that there are I populations, that $n^{(i)}$ individuals are sampled from the i^{th} population, and that $p^{(i)}$ variables are measured from each individual sampled from the i^{th} population, $i = 1, 2, \cdots, I$. Let $\mathbf{w}_j^{(i)}$ denote the $p^{(i)} \times 1$ observation from the j^{th} individual, $j = 1, 2, \cdots, n^{(i)}$, from the i^{th} population. The general latent variable model assumes that $\mathbf{w}_j^{(i)}$ is a linear function of some underlying unobservable factor vector and error vector. Particular structures of the coefficients and the factor and error covariance structure produce models with particular structures for the moments of $\mathbf{w}_j^{(i)}$. For technical reasons, it is useful to consider an augmented vector $\mathbf{z}_j^{(i)} = (\mathbf{w}_j^{(i)\prime}, 1)'$ and express the model in terms of $\mathbf{z}_j^{(i)}$. A general structural equation model assumes that the vector $\mathbf{z}_j^{(i)}$ can be expressed as a linear combination of latent vector variables $\mathbf{f}_{\ell j}^{(i)}$ of dimension $q_\ell^{(i)} \times 1$ premultiplied by loading matrices $\mathbf{B}_\ell^{(i)}$ of dimension $(p^{(i)}+1) \times q_\ell^{(i)}$; $\ell = 0, 1, 2, \cdots, L^{(i)}$,

$$\mathbf{z}_j^{(i)} = \mathbf{B}_0^{(i)}\mathbf{f}_{0j}^{(i)} + \mathbf{B}_1^{(i)}\mathbf{f}_{1j}^{(i)} + \cdots + \mathbf{B}_{L^{(i)}}^{(i)}\mathbf{f}_{L^{(i)}j}^{(i)}, \tag{3.4}$$

where the $\mathbf{f}_{0j}^{(i)}$ are augmented vectors $\mathbf{f}_{0j}^{(i)} = (\mathbf{g}_{0j}^{(i)\prime},\ 1)'$. In this paper, four special cases of (3.4) are considered depending on the distributional assumption for $\mathbf{f}_{\ell j}^{(i)}$.

(1) The model FNR (Fixed, Normal, Random (non-normal)) assumes that the $\mathbf{f}_{0j}^{(i)}$ are fixed constants satisfying

$$\lim_{n^{(i)}\to\infty} \boldsymbol{\Phi}_{00}^{(i)}(n^{(i)}) = \lim_{n^{(i)}\to\infty} \frac{1}{n^{(i)}} \sum_{j=1}^{n^{(i)}} \mathbf{f}_{0j}^{(i)}\mathbf{f}_{0j}^{(i)\prime} = \boldsymbol{\Phi}_{00}^{(i)o} \tag{3.5}$$

for some positive definite $\boldsymbol{\Phi}_{00}^{(i)o}$, that the $\mathbf{f}_{\ell j}^{(i)}$ $\ell = 1, 2, \cdots, L^{(i)}$ are assumed to be independently identically distributed with mean zero and positive definite covariance matrices $\boldsymbol{\Phi}_{\ell\ell}^{(i)}$, and that the $\mathbf{f}_{1j}^{(i)}$ are normally distributed.

(2) The model FN (Fixed, Normal) is the model FNR with normally distributed $\mathbf{f}_{\ell j}^{(i)}$, $\ell = 2, 3, \cdots, L^{(i)}$.

(3) The model NR (Normal, Random) assumes that the $\mathbf{f}_{\ell j}^{(i)}$, $\ell = 0, 1, \cdots, L^{(i)}$, are independently identically distributed. The $\mathbf{f}_{0j}^{(i)}$ have mean non-zero and positive definite second-order moment matrices $\boldsymbol{\Phi}_{00}^{(i)}$, the $\mathbf{f}_{\ell j}^{(i)}$ $\ell = 1, 2, \cdots, L^{(i)}$ have mean zero and positive definite covariance matrices $\boldsymbol{\Phi}_{\ell\ell}^{(i)}$, and the $\mathbf{f}_{1j}^{(i)}$ are normally distributed.

(4) The model N (Normal) is the model NR with normally distributed $\mathbf{f}_{\ell j}^{(i)}$ $\ell = 0, 1, \cdots, L^{(i)}$.

For all four models, it is assumed that the $\mathbf{B}_\ell^{(i)}$, $\ell = 0, 1, \cdots, L^{(i)}$ and $\mathbf{\Phi}_{11}^{(i)}$ are functions of a vector $\boldsymbol{\tau}$ of dimension $r \times 1$ for all $i = 1, 2, \cdots, I$. That is, $\mathbf{B}_\ell^{(i)} = \mathbf{B}_\ell^{(i)}(\boldsymbol{\tau})$ and $\mathbf{\Phi}_{11}^{(i)} = \mathbf{\Phi}_{11}^{(i)}(\boldsymbol{\tau})$. Note that for $\ell = 1, 2, \cdots, L^{(i)}$ $\mathbf{\Phi}_{\ell\ell}^{(i)}$ is the covariance matrix of $\mathbf{f}_{\ell j}^{(i)}$, and that $\mathbf{\Phi}_{00}^{(i)}$ for models NR and N contains the first two moments of $\mathbf{g}_{oj}^{(i)}$ where $\mathbf{f}_{0j}^{(i)} = (\mathbf{g}_{0j}^{(i)\prime},\ 1)'$.

Example: Consider the structural equations (3.1), (3.2), and (3.3) in I populations, and assume that $\boldsymbol{\xi}_j^{(i)}$ are fixed, that $\boldsymbol{\zeta}_j^{(i)}$, $\boldsymbol{\epsilon}_j^{(i)}$, and $\boldsymbol{\delta}_j^{(i)}$ are independent random vectors with mean zero and diagonal covariance matrices, and that the $\boldsymbol{\zeta}_j^{(i)}$ follow normal distribution. The following equation shows that this model can be written in the form of the model FNR:

$$\mathbf{z}_j^{(i)} = \begin{pmatrix} \mathbf{y}_j^{(i)} \\ \mathbf{x}_j^{(i)} \\ 1 \end{pmatrix} = \begin{pmatrix} \mathbf{A}_1^{(i)} & \mathbf{A}_2^{(i)} \\ \mathbf{\Lambda}_x^{(i)} & \boldsymbol{\tau}_x^{(i)} \\ \mathbf{0} & 1 \end{pmatrix} \begin{pmatrix} \boldsymbol{\xi}_j^{(i)} \\ 1 \end{pmatrix} + \begin{pmatrix} \mathbf{\Lambda}_y^{(i)}(I - \mathbf{B}^{(i)})^{-1} \\ \mathbf{0} \end{pmatrix} \boldsymbol{\zeta}_j^{(i)}$$

$$+ \begin{pmatrix} \mathbf{I}_{\bullet 1} \\ \mathbf{0} \end{pmatrix} \epsilon_{1j}^{(i)} + \cdots + \begin{pmatrix} \mathbf{I}_{\bullet p_y^{(i)}} \\ \mathbf{0} \end{pmatrix} \epsilon_{p_y^{(i)} j}^{(i)} + \begin{pmatrix} \mathbf{0} \\ \mathbf{I}_{\bullet 1} \\ 0 \end{pmatrix} \delta_{1j}^{(i)} + \cdots + \begin{pmatrix} \mathbf{0} \\ \mathbf{I}_{\bullet p_x^{(i)}} \\ 0 \end{pmatrix} \delta_{p_x^{(i)} j}^{(i)},$$

where $\mathbf{A}_1^{(i)} = \mathbf{\Lambda}_y^{(i)}(I - \mathbf{B}^{(i)})\mathbf{\Gamma}^{(i)}$, $\mathbf{A}_2^{(i)} = \boldsymbol{\tau}_y^{(i)} + \mathbf{\Lambda}_y^{(i)}(I - \mathbf{B}^{(i)})^{-1}\boldsymbol{\alpha}^{(i)}$, $p_x^{(i)}$ and $p_y^{(i)}$ are the dimensions of the vectors $\mathbf{x}_j^{(i)}$ and $\mathbf{y}_j^{(i)}$ respectively, $\epsilon_{kj}^{(i)}$ and $\delta_{kj}^{(i)}$ are the k^{th} components of $\boldsymbol{\epsilon}_j^{(i)}$ and $\boldsymbol{\delta}_j^{(i)}$, and $\mathbf{I}_{\bullet k}$ is the k^{th} column of the identity matrix $\mathbf{I}$. This is the model FNR where $\mathbf{f}_{0j}^{(i)} = (\boldsymbol{\xi}_j^{(i)\prime}, 1)'$, $\mathbf{f}_{1j}^{(i)} = \boldsymbol{\zeta}_j^{(i)}$, $\mathbf{f}_{\ell j}^{(i)} = \epsilon_{\ell-1,j}^{(i)}$ for $\ell = 2, 3, \cdots, p_y^{(i)} + 1$, and $\mathbf{f}_{\ell j}^{(i)} = \delta_{\ell - p_y^{(i)} - 1, j}^{(i)}$ for $\ell = p_y^{(i)} + 2, \cdots, p_x^{(i)} + p_y^{(i)} + 1$, and $\boldsymbol{\tau}$ consists of the unknown elements of $\mathbf{B}^{(i)}$, $\mathbf{\Gamma}^{(i)}$, $\mathbf{\Lambda}_y^{(i)}$, $\mathbf{\Lambda}_x^{(i)}$, $\boldsymbol{\tau}_x^{(i)}$, $\boldsymbol{\tau}_y^{(i)}$, $\boldsymbol{\alpha}^{(i)}$, and $Var\{\boldsymbol{\zeta}_j^{(i)}\}$. □

Our interest is in the asymptotic robustness of the maximum likelihood estimator under model N when the true model is in fact model FNR, FN, or NR. The maximum likelihood estimator under model N can be readily computed using the existing packages. Under model N, the unknown parameters can be written together in a vector $\boldsymbol{\theta}$ as

$$\boldsymbol{\theta} = (\boldsymbol{\tau}', \boldsymbol{\varphi}', \boldsymbol{\psi}')',$$

where

$$\begin{aligned}\boldsymbol{\varphi}' &= (vech'\boldsymbol{\Phi}_{22}^{(1)}, \cdots, vech'\boldsymbol{\Phi}_{L^{(1)}L^{(1)}}^{(1)}, \cdots, vech'\boldsymbol{\Phi}_{22}^{(I)}, \cdots, vech'\boldsymbol{\Phi}_{L^{(I)}L^{(I)}}^{(I)}),\\ \boldsymbol{\psi}' &= (vech'\boldsymbol{\Phi}_{00}^{(1)}, \cdots, vech'\boldsymbol{\Phi}_{00}^{(I)}).\end{aligned}$$

In this paper the *vec* and *vech* notation is used. For definitions and properties of the *vec* and *vech* operators, see, e.g., Fuller [11]. The dimension of $\boldsymbol{\theta}$ is

$$k = r + \sum_{i=1}^{I}[\sum_{\substack{\ell=0\\ \ell\neq 1}}^{L^{(i)}} \frac{q_\ell^{(i)}(q_\ell^{(i)}+1)}{2} - 1],$$

where the subtraction of 1 is due to the structure of $\boldsymbol{\Phi}_{00}^{(i)}$ with the known value of 1 in the lower right corner. It is assumed for identification that

$$k \leq \sum_{i=1}^{I}[\frac{p^{(i)}(p^{(i)}+1)}{2} + p^{(i)}].$$

To compute the maximum likelihood estimator under model N, let

$$\mathbf{C}^{(i)} = \frac{1}{n^{(i)}}\sum_{j=1}^{n^{(i)}} \mathbf{z}_j^{(i)}\mathbf{z}_j^{(i)\prime}$$

be the sample augmented moment matrices of observation vectors. For the use of augmented moment matrices, see, e.g, Jöreskog and Sörbom [13], Meredith and Tisak [15], and Satorra [17, 18, 19]. Note that the information contained in $\mathbf{C}^{(i)}$ is equivalent to that given by the first two sample moments. Let $\boldsymbol{\Gamma}^{(i)}(\boldsymbol{\theta}) = E\{\mathbf{C}^{(i)}\}$ under model N, which is a function of $\boldsymbol{\theta}$. Also, define

$$\begin{aligned} \mathbf{c} &= (vech'\mathbf{C}^{(1)}, \cdots, vech'\mathbf{C}^{(I)})', \\ \boldsymbol{\gamma}(\boldsymbol{\theta}) &= (vech'\boldsymbol{\Gamma}^{(1)}(\boldsymbol{\theta}), \cdots, vech'\boldsymbol{\Gamma}^{(I)}(\boldsymbol{\theta}))'. \end{aligned}$$

Then, it can be shown that the maximum likelihood estimator $\hat{\boldsymbol{\theta}}$ under model N can be obtained by minimizing the function

$$L[\boldsymbol{\gamma}(\boldsymbol{\theta}), \mathbf{c}] = \sum_{i=1}^{I} \frac{n^{(i)}}{n} [\log |\boldsymbol{\Gamma}^{(i)}(\boldsymbol{\theta})| + tr\{\mathbf{C}^{(i)}\boldsymbol{\Gamma}^{(i)-1}(\boldsymbol{\theta})\} - \log |\mathbf{C}^{(i)}| - (p^{(i)} + 1)] \quad (3.6)$$

over the parameter space $\boldsymbol{\Theta} = \{\boldsymbol{\theta} \mid \boldsymbol{\tau} \in \boldsymbol{\Theta}_{\boldsymbol{\tau}} \subseteq \Re^k$ and $\boldsymbol{\Phi}_{\ell\ell}^{(i)}$ are symmetric nonnegative definite, $\ell = 0, 1, \cdots, L^{(i)}\}$, where $n = \sum_{i=1}^{I} n^{(i)}$. The lower right corner of $\boldsymbol{\Phi}_{00}^{(i)}$ is known to be 1, and is not estimated. Note that $L[\boldsymbol{\gamma}(\boldsymbol{\theta}), \mathbf{c}]$ is a discrepancy function. That is, 1) $L(\boldsymbol{\gamma}, \mathbf{c}) \geq 0$, 2) $L(\boldsymbol{\gamma}, \mathbf{c}) = 0 \Leftrightarrow \boldsymbol{\gamma} = \mathbf{c}$, and 3) $L(\boldsymbol{\gamma}, \mathbf{c})$ is a twice continuously differentiable function of $\boldsymbol{\gamma}$ and $\mathbf{c}$. Although $\hat{\boldsymbol{\theta}}$ can be computed for any data, the $\boldsymbol{\psi}$ part of $\boldsymbol{\theta}$ does not have a corresponding model parameter in models FNR and FN where $\mathbf{f}_{0j}^{(i)}$ are fixed. To handle the fixed $\mathbf{f}_{0j}^{(i)}$ problem, we must define the true values in a slightly different way. Let $\boldsymbol{\tau}_o$ and $\boldsymbol{\varphi}_o$ be the true parameter values of $\boldsymbol{\tau}$ and $\boldsymbol{\varphi}$ under all four models. Let $\boldsymbol{\psi}_o$ denote the true parameter value in models NR and N, and denote

$$(vech'\boldsymbol{\Phi}_{00}^{(1)}(n^{(1)}), \cdots, vech'\boldsymbol{\Phi}_{00}^{(I)}(n^{(I)})) \quad (3.7)$$

in models FNR and FN. Also, define $\boldsymbol{\psi}_\infty$ to be $\boldsymbol{\psi}_o$ in models NR and N and the corresponding limiting value of (3.7), defining in(3.5), under models FNR and FN. Also, let $\boldsymbol{\theta}_o = (\boldsymbol{\tau}_o', \boldsymbol{\varphi}_o', \boldsymbol{\psi}_o')'$, $\boldsymbol{\theta}_\infty = (\boldsymbol{\tau}_o', \boldsymbol{\varphi}_o', \boldsymbol{\psi}_\infty')'$, and $n_m = min\{n^{(1)}, n^{(2)}, \cdots, n^{(I)}\}$. We also write, for a $p \times p$ symmetric $\mathbf{A}$, $vec\mathbf{A} = \mathbf{K}_p vech\mathbf{A}$ and $vech\mathbf{A} = \mathbf{K}_p^+ vec\mathbf{A}$, where $\mathbf{K}_p^+ = (\mathbf{K}_p'\mathbf{K}_p)^{-1}\mathbf{K}_p'$. Denote, for $i = 1, 2, \cdots, I$,

$$\mathbf{\Omega}^{(i)} = 2\mathbf{K}_{p(i)}^+[\mathbf{\Gamma}^{(i)}(\boldsymbol{\theta}_\infty) \otimes \mathbf{\Gamma}^{(i)}(\boldsymbol{\theta}_\infty)]\mathbf{K}_{p(i)}^{+\prime},$$

and it follows (see, e.g., Fuller [11])

$$\mathbf{\Omega}^{(i)-1} = \frac{1}{2}\mathbf{K}_{p(i)}'[\mathbf{\Gamma}^{(i)-1}(\boldsymbol{\theta}_\infty) \otimes \mathbf{\Gamma}^{(i)-1}(\boldsymbol{\theta}_\infty)]\mathbf{K}_{p(i)}.$$

Define

$$\mathbf{\Omega}^{-1} = \text{blockdiag}(r^{(1)}\mathbf{\Omega}^{(1)-1}, \cdots, r^{(I)}\mathbf{\Omega}^{(I)-1})$$

where $r^{(i)} = \lim_{n_m \to \infty}(n^{(i)}/n) \quad i = 1, 2, \cdots, I$. Also, let

$$\mathbf{F} = [\mathbf{F}^{(1)\prime}, \cdots, \mathbf{F}^{(I)\prime}]',$$

where

$$\mathbf{F}^{(i)} = \left.\frac{\partial vech \mathbf{\Gamma}^{(i)}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_\infty}.$$

3.4 Asymptotic Robustness

This section presents only some summary results without proofs. More detailed results and discussions as well as proofs are given elsewhere. The following regularity condition is assumed throughout this section.

Assumption 1

i) $\boldsymbol{\theta}_\infty$ *is an interior point of the parameter space* $\mathbf{\Theta}$.

ii) For any $\varepsilon > 0$ *there exists* $\delta > 0$ *such that if* $\| \gamma(\boldsymbol{\theta}) - \gamma(\boldsymbol{\theta}_\infty) \| < \delta$ *then* $\| \boldsymbol{\theta} - \boldsymbol{\theta}_\infty \| < \varepsilon$ *where* $\| \boldsymbol{\theta} \| = \sqrt{\boldsymbol{\theta}'\boldsymbol{\theta}}$.

iii) $\lim_{n_m \to \infty}(n^{(i)}/n) = r^{(i)} > 0 \;\; i = 1, 2, \cdots, I.$

iv) The $\mathbf{B}_\ell^{(i)}(\boldsymbol{\tau})$ *and* $\boldsymbol{\Phi}_{11}^{(i)}(\boldsymbol{\tau})$, $\ell = 0, 1, \cdots, L^{(i)}; i = 1, 2, \cdots, I$ *are twice continuously differentiable in a neighborhood of* $\boldsymbol{\tau}_o$.

v) $\mathbf{F}(\boldsymbol{\theta}_\infty)$ *has full column rank.*

The first theorem gives the limiting distribution of $\hat{\boldsymbol{\theta}}$ defined in (3.6) when the true model is the model N, NR, FN, or FNR.

Theorem 1 *Let Assumption 1 hold.*

a) Under the model N, as $n_m \to \infty$,

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_o) \xrightarrow{L} \mathbf{N}(\mathbf{0}, \mathbf{V}_N),$$

where

$$\mathbf{V}_N = (\mathbf{F}'\boldsymbol{\Omega}^{-1}\mathbf{F})^{-1}.$$

b) Let the model NR hold. If the $\mathbf{f}_{\ell j}^{(i)}$ *for all* $\ell \neq 1$ *have finite fourth moments, then, as* $n_m \to \infty$,

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_o) \xrightarrow{L} \mathbf{N}(\mathbf{0}, \mathbf{V}_{NR}),$$

where

$$\mathbf{V}_{NR} = (\mathbf{F}'\boldsymbol{\Omega}^{-1}\mathbf{F})^{-1}\mathbf{F}'\boldsymbol{\Omega}^{-1}\boldsymbol{\Upsilon}_{NR}\boldsymbol{\Omega}^{-1}\mathbf{F}(\mathbf{F}'\boldsymbol{\Omega}^{-1}\mathbf{F})^{-1},$$

$$\boldsymbol{\Upsilon}_{NR} = blockdiag(r^{(1)-1}\boldsymbol{\Upsilon}_{NR}^{(1)}, \cdots, r^{(I)-1}\boldsymbol{\Upsilon}_{NR}^{(I)}),$$

$$\boldsymbol{\Upsilon}_{NR}^{(i)} = Var\{vech(\mathbf{z}_j^{(i)}\mathbf{z}_j^{(i)\prime})\}.$$

c) Under the model FN, as $n_m \to \infty$.

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_o) \xrightarrow{L} \mathbf{N}(\mathbf{0}, \mathbf{V}_{FN}),$$

where $\mathbf{V}_{FN}$ *has the same form as* $\mathbf{V}_{NR}$ *with* $\boldsymbol{\Upsilon}_{NR}$ *replaced by* $\boldsymbol{\Upsilon}_{FN}$*,* $\boldsymbol{\Upsilon}_{FN}$ *has the same form as* $\boldsymbol{\Upsilon}_{NR}$ *with* $\boldsymbol{\Upsilon}_{NR}^{(i)}$ *replaced by* $\boldsymbol{\Upsilon}_{FN}^{(i)}$*, and*

$$\boldsymbol{\Upsilon}_{FN}^{(i)} = \boldsymbol{\Omega}^{(i)} - 2\mathbf{K}_{p^{(i)}}^{+}[(\mathbf{B}_0^{(i)}\boldsymbol{\Phi}_{00}^{(i)o}\mathbf{B}_0^{(i)\prime}) \otimes (\mathbf{B}_0^{(i)}\boldsymbol{\Phi}_{00}^{(i)o}\mathbf{B}_0^{(i)\prime})]\mathbf{K}_{p^{(i)}}^{+\prime}.$$

d) Let the model FNR hold. If the $\mathbf{f}_{\ell j}^{(i)}$ *for all* $\ell \neq 1$ *have finite fourth moments, then, as* $n_m \to \infty$*,*

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_o) \xrightarrow{L} \mathbf{N}(\mathbf{0}, \mathbf{V}_{FNR}),$$

where $\mathbf{V}_{FNR}$ *has the same form as* $\mathbf{V}_{NR}$ *with* $\boldsymbol{\Upsilon}_{NR}$ *replaced by* $\boldsymbol{\Upsilon}_{FNR}$*,* $\boldsymbol{\Upsilon}_{FNR}$ *has the same form as* $\boldsymbol{\Upsilon}_{NR}$ *with* $\boldsymbol{\Upsilon}_{NR}^{(i)}$ *replaced by* $\boldsymbol{\Upsilon}_{FNR}^{(i)}$*, and*

$$\boldsymbol{\Upsilon}_{FNR}^{(i)} = \boldsymbol{\Upsilon}_{FN}^{(i)} + \sum_{\ell=2}^{L^{(i)}}(\mathbf{H}_\ell^{(i)} + \mathbf{G}_\ell^{(i)} + \mathbf{G}_\ell^{(i)\prime})$$

where

$$\begin{aligned}
\mathbf{H}_\ell^{(i)} &= \mathbf{K}_{p^{(i)}}^{+}(\mathbf{B}_\ell^{(i)} \otimes \mathbf{B}_\ell^{(i)})\mathbf{K}_{q_\ell^{(i)}}[Var\{vech(\mathbf{f}_{\ell j}^{(i)}\mathbf{f}_{\ell j}^{(i)\prime})\} \\
&\quad -2\mathbf{K}_{q_\ell^{(i)}}^{+}(\boldsymbol{\Phi}_{\ell\ell}^{(i)o} \otimes \boldsymbol{\Phi}_{\ell\ell}^{(i)o})\mathbf{K}_{q_\ell^{(i)}}^{+\prime}]\mathbf{K}_{q_\ell^{(i)}}^{\prime}(\mathbf{B}_\ell^{(i)} \otimes \mathbf{B}_\ell^{(i)})^{\prime}\mathbf{K}_{p^{(i)}}^{+\prime}, \\
\mathbf{G}_\ell^{(i)} &= 2\mathbf{K}_{p^{(i)}}^{+}(\mathbf{B}_\ell^{(i)} \otimes \mathbf{B}_\ell^{(i)})\mathbf{K}_{q_\ell^{(i)}}E\{vech(\mathbf{f}_{\ell j}^{(i)}\mathbf{f}_{\ell j}^{(i)\prime})\mathbf{f}_{\ell j}^{(i)\prime}\} \times (\mathbf{B}_\ell^{(i)} \otimes \boldsymbol{\mu}^{(i)})^{\prime}\mathbf{K}_{p^{(i)}}^{+\prime}], \\
\boldsymbol{\mu}^{(i)} &= \lim_{n^{(i)}\to\infty} \frac{1}{n^{(i)}}\sum_{j=1}^{n^{(i)}} \mathbf{z}_j^{(i)} = \lim_{n^{(i)}\to\infty} \bar{\mathbf{z}}^{(i)},
\end{aligned}$$

and $\boldsymbol{\Phi}_{\ell\ell}^{(i)o}$ *are the true values of* $\boldsymbol{\Phi}_{\ell\ell}^{(i)}$*,* $\ell = 1, 2, \cdots, L^{(i)}$*.*

The results of Theorem 1 provide the limiting distribution of the whole parameter estimate vector $\hat{\boldsymbol{\theta}}$, and require the existence of the fourth moments of the random variables. However, the asymptotic robustness, i.e., the equality of the limiting covariance matrix over the four models, can be shown for the $\boldsymbol{\tau}$ part of $\hat{\boldsymbol{\theta}}$ with only the existence of second moments.

Theorem 2 *Let Assumption 1 hold. Then, under each of models N, NR, FN, FNR, as $n \to \infty$,*

$$\sqrt{n}(\hat{\boldsymbol{\tau}} - \boldsymbol{\tau}_o) \xrightarrow{L} \mathbf{N}(\mathbf{0}, \mathbf{V}^{\boldsymbol{\tau}}), \tag{3.8}$$

where $\mathbf{V}^{\boldsymbol{\tau}}$ is the part of $\mathbf{V}_N$ corresponding to $\boldsymbol{\tau}$.

Thus, the standard error estimate for $\boldsymbol{\tau}$ computed under model N is also valid for the three other models. Some asymptotic robustness results for $\hat{\boldsymbol{\varphi}}$ and $\hat{\boldsymbol{\psi}}$ can also be derived, but are not presented here. It should be pointed out that Satorra [20] has a result similar to Theorem 2.

For possibly dependent populations (e.g., repeated measures), we have the following result.

Theorem 3 *Let Assumption 1 hold. Consider model NR where $\mathbf{f}_{0j}^{(i)}$ are independent of all $\mathbf{f}_{\ell j}^{(i)}$, $\ell = 1, 2, \cdots, L^{(i)}$, but may be dependent over $i = 1, 2, \cdots, I$. Then, (3.8) holds.*

Note that in Theorem 3 the assumption on $\mathbf{f}_{0j}^{(i)}$ is very weak allowing any kind of dependency over populations. Hence, by including any latent variables with possible correlation over populations in one vector $\mathbf{f}_{0j}^{(i)}$, the asymptotic inference on $\boldsymbol{\tau}$ based on the assumption of independent normal populations is still valid for a large class of dependent non-normal populations. An example of such a situation is the use of the standard normal-independence based packages for analyzing non-normal samples obtained by repeatedly measuring individuals longitudinally where not all individuals appear at all time points.

This paper concentrated on the maximum likelihood estimator $\hat{\boldsymbol{\theta}}$, for model N by minimizing $L[\boldsymbol{\gamma}, \mathbf{c}]$ defining in (3.6). There are other discrepancy functions that can be used to define alternative estimators of $\boldsymbol{\theta}$. Results similar to Theorems 1, 2, and 3 hold for some estimators other than $\hat{\boldsymbol{\theta}}$, but are not given here. In addition, the asymptotic robustness results for the goodness-of-fit test statistics can also be derived, but are not discussed in this paper. It should be emphasized that the condition of unrestricted $\boldsymbol{\Phi}_{\ell\ell}^{(i)}$, $\ell \neq 1$, included in all four models is crucial for the asymptotic robustness results in Theorem 2 and 3. Our simulation, which is not reported here, indicated that without this condition the results of Theorem 2 and 3 are severely violated in finite samples.

3.5 Monte Carlo Study

To illustrate the usefulness of the asymptotic results of this paper, simple regression with errors in variables is considered in $I = 2$ samples,

$$\begin{cases} y_j^{(i)} &= \tau_y + \lambda_y \eta_j^{(i)} + \varepsilon_j^{(i)} \\ x_j^{(i)} &= \qquad\quad \eta_j^{(i)} + \delta_j^{(i)} \end{cases}$$

for $i = 1, 2$; $j = 1, 2, \cdots, 1000$. Such a set of two samples was generated 10,000 times. The $\eta_j^{(i)}$ are fixed latent variables, and $\varepsilon_j^{(i)}$ and $\delta_j^{(i)}$ are generated as non-normal variables. This is a special case of the model FNR. Two sets of 1000 uniform random variables were generated as $\eta_j^{(1)}$ and $\eta_j^{(2)}$ with the sample correlation between $\eta_j^{(1)}$ and $\eta_j^{(2)}$ being 0.7. As fixed variables, these two sets of $\eta_j^{(i)}$ values were kept unchanged over the 10,000 replications. The $\varepsilon_j^{(i)}$ and $\delta_j^{(i)}$ are independent adjusted chi-square random variables with one degree of freedom, and with mean zero and Var($\varepsilon_j^{(i)}$)= $\phi_{\varepsilon\varepsilon}^{(i)}$, Var($\delta_j^{(i)}$)= $\phi_{\delta\delta}^{(i)}$, for i=1,2. Note that the intercept τ_y and the slope λ_y are common for the two populations, and that the variances of the $\varepsilon_j^{(i)}$ and $\delta_j^{(i)}$ are unrestricted.

In this example $\boldsymbol{\theta} = (\boldsymbol{\tau}', \boldsymbol{\varphi}', \boldsymbol{\psi}')'$, where $\boldsymbol{\tau} = (\tau_y, \lambda_y)'$,

$$\boldsymbol{\varphi} = (\phi_{\epsilon\epsilon}^{(1)},\ \phi_{\delta\delta}^{(1)},\ \phi_{\epsilon\epsilon}^{(2)},\ \phi_{\delta\delta}^{(2)})',$$

and $\boldsymbol{\psi}$ contains the first two sample moments of $\eta_j^{(i)}$'s. The true parameter values are $\boldsymbol{\tau}_o = (1,\ 2)$ and $\boldsymbol{\varphi}_o = (0.2,\ 0.2,\ 0.3,\ 0.15)'$. The true sample moments of $\eta_j^{(i)}$'s are

$$\bar{\eta}^{(1)} = \frac{1}{1000}\sum_{j=1}^{1000} \eta_j^{(1)} = 2,$$

$$s_{\eta\eta}^{(1)} = \frac{1}{1000}\sum_{j=1}^{1000} \eta_j^{(1)2} = 4.9,$$

$\bar{\eta}^{(2)} = 1$, and $s_{\eta\eta}^{(2)} = 2.5$.

Table 3.1 summarizes the results. The second column of Table 3.1 gives the Monte Carlo (MC) standard errors and the last four columns give the asymptotic standard errors computed under the assumptions of the models FNR, FN, NR, and N. By the results in Section 3.4 and elsewhere, some asymptotic standard errors are known to be equal over models. These are indicated by bold-face. Also, these bold-face asymptotic standard errors are supposed to be approximations of the exact Monte Carlo standard errors. These approximations seem to be very good. Clearly, the asymptotic standard errors not supported by the theory (indicated by non-bold-face) are poor estimates of the exact standard errors, and are not useful in practice.

Table 3.1: Monte Carlo (MC) and asymptotic standard errors for four models (FNR, FN, NR, and N)

$\hat{\boldsymbol{\theta}}$		MC	FNR	FN	NR	N
$\hat{\boldsymbol{\tau}}$	$\hat{\tau}_y$	.065	**.066**	**.066**	**.066**	**.066**
	$\hat{\lambda}_y$	.044	**.043**	**.043**	**.043**	**.043**
$\hat{\boldsymbol{\varphi}}$	$\hat{\phi}^{(1)}_{\epsilon\epsilon}$	.109	**.111**	.087	**.111**	.087
	$\hat{\phi}^{(1)}_{\delta\delta}$	.028	**.028**	.023	**.028**	.023
	$\hat{\phi}^{(2)}_{\epsilon\epsilon}$	.168	**.171**	.138	**.171**	.138
	$\hat{\phi}^{(2)}_{\delta\delta}$	.047	**.048**	.039	**.048**	.039
$\hat{\boldsymbol{\psi}}$	$\hat{\bar{\eta}}^{(1)}$	.014	**.014**	**.014**	.035	.035
	$\hat{s}^{(1)}_{\eta\eta}$	.072	**.073**	**.073**	.150	.153
	$\hat{\bar{\eta}}^{(2)}$	.012	**.012**	**.012**	.041	.041
	$\hat{s}^{(2)}_{\eta\eta}$	.034	**.033**	**.033**	.097	.108

3.6 References

[1] Amemiya, Y., Fuller W.A. and Pantula, S.G. (1987). The asymptotic distributions of some estimators for a factor analysis model. *Journal of Multivariate Analysis*, **22**, 51-64.

[2] Amemiya, Y. and Anderson, T.W. (1990). Asymptotic chi-square tests for a large class of factor analysis models. *The Annals of Statistics*, **18**, 1453-1463.

[3] Anderson, T.W. (1987). Multivariate linear relations. *Proceedings of the Second International Conference in Statistics,* edited by Pukkila T. and Puntanen S., University of Tampere, Finland. 9-36.

[4] Anderson, T.W. (1989). Linear latent variable models and covariance structures. *Journal of Econometrics*, **41**, 91-119.

[5] Anderson, T.W. and Amemiya, Y.(1988). The asymptotic normal distribution of estimators in factor analysis under general conditions. *The Annals of Statistics,* **16**, 759-771.

[6] Anderson, T.W. and Rubin, H. (1956). Statistical inference in factor analysis. *Proceedings of Third Berkeley Symposium.* Univ. of California Press, Berkeley. 111-150.

[7] Bentler, P.M. (1989). *EQS Structural Equations Program Manual.* BMDP Statistical Software, Inc., Los Angeles.

[8] Bollen, P.A. (1989). *Structural Equations with Latent Variables.* John Wiley and Sons, New York.

[9] Browne, M.W. (1988). Asymptotic robustness of normal theory methods for the analysis of latent variables. *Statistical Analysis of Measurement Errors and Applications*, edited by Brown, P.J. and Fuller, W.A.: American Mathematical Society.

[10] Browne, M.W. and Shapiro, A. (1988). Robustness of normal theory methods in the analysis of linear latent variate models. *British Journal of Mathematical and Statistical Society*, **41**, 193-208.

[11] Fuller, W.A. (1987). *Measurement Error Models.* John Wiley and Sons, New York.

[12] Jõreskog, K. (1971). Simultaneous factor analysis in several populations. *Psychometrika.* **47**, 297-308.

[13] Jõreskog, K. and Sõrbom, D. (1989). *LISREL 7; A Guide to the Program and Applications.* 2nd ed., SPSS INC., Chicago.

[14] Lee, S.Y. and Tsui, K.L. (1982). Covariance structure analysis in several populations. *Psychometrika*, **55**, 107-122.

[15] Meredith, W. and Tisak, A. (1990). Latent curve analysis. *Psychometrika*, **55**, 107-122.

[16] Muthén, B. (1989). Multiple group structural modeling with non-normal continuous variables. *British Journal of Mathematical and Statistical Psychology*, **42**, 55-62.

[17] Satorra, A. (1992). Asymptotic robust inferences in the analysis of mean and covariance structures. *Sociological Methodology,* edited by Marsden, P.V., 249-278.

[18] Satorra, A. (1993). Multi-sample analysis of moment-structures: Asymptotic validity of inferences based on second-order moments. *Statistical Modeling and Latent Variables,* edited by Haagen, K., Bartholomew, D.J. and Deistler, M.. Amsterdam: Elsevier. Forthcoming.

[19] Satorra, A. (1993). Asymptotic robust inferences multi-sample analysis of augmented-moment structures. *Multivariate Analysis: Future Directions 2*, edited by Cuadras, C.M. and Rao, C.R., 211-229.

[20] Satorra, A. (1994). On asymptotic robustness in multiple-group analysis of multivariate relations. Paper presented at *Latent Variable Modeling with Applications to Causality*, March 19-20, Los Angeles.

4. FACTOR ANALYSIS OF LONGITUDINAL DATA

A paper to be submitted to the Journal of Multivariate Analysis [1]

Savas Papadopoulos and Yasuo Amemiya

4.1 Abstract

Factor analysis of multivariate panel data is discussed, where measurements are taken from individuals at several occasions. Unbalanced cases, in which some individuals do not appear at all occasions and the number of measured individuals may change from one occasion to another, are considered. For such cases, a relatively simple method based on an incorrect likelihood is suggested. The method can be implemented easily utilizing existing computer packages, and is shown to have various advantages over the maximum normal likelihood estimation and the time series modeling. The inference procedures are asymptotically valid for non-normal data allowing any time trend. For the balanced normal case, the efficiency of the method is shown to be nearly as high as that of the maximum likelihood estimation.

[1] *Key words:* Latent variable modeling, non-normal data, multivariate repeated measures.

4.2 Introduction

In medical and social studies, individuals are often observed over time. Data collected from such studies are called longitudinal data. Longitudinal data are known also as panel data or repeated measures. In many longitudinal studies, some individuals are missed or added at specific times. We consider a situation where several variables are measured from an individual at each occasion, and where the relationships among the observed variables are explained in terms of relatively few underlying factors. The factors or latent variables represent unobservable characteristics of individuals which are correlated over time. The number of the observed variables and the number of the factors may change over time. For such a general longitudinal factor analysis problem, this paper proposes a simple but widely-applicable approach for model fitting and checking, and for parameter estimation.

For general introduction to factor analysis and longitudinal factor analysis models, see, e.g., Bentler (1989), Bollen (1989), Jõreskog and Sõrbom (1989), and Basilevsky (1994). In the special case of balanced data where all the individuals appear at all the occasions, the longitudinal factor analysis model can be written in the form of the regular factor analysis model for all observed variables. See Anderson (1987, 1989). In some situations, the factors are assumed to follow an autoregressive moving average process, and time series modeling is applied. See, Molenaar (1985, 1992).

The treatment of unbalanced data, where individuals are missed or added over time, can be considered as a missing observation problem. However, the standard methods for missing data do not lead to useful or efficient procedures in practical longitudinal factor analysis problems. For example, in the listwise deletion method using only those individuals appearing at all the occasions, the sample size may be dramatically reduced, and the estimators may not be efficient. For the pairwise deletion method using the sample variances and covariances computed with all possible pairwise combinations, the resulting sample covariance matrix may not be positive definite, and the computed standard errors and chi-square tests are incorrect. Also, the imputation procedures are often time consuming, and the development of appropriate procedures under imputation can be difficult. Another approach is the so-called multi-sample analysis, where groups of individuals are formed based on the occasion appearance patterns. In applying this method, the number of groups may be very large (up to $2^T - 1$ for T occasions), and some groups may contain only a few individuals. See, e.g., Werts, Rock and Grany (1979), Baker and Fulker (1983), and Allison (1987).

If the longitudinal factor analysis model was completely specified, the full likelihood approach would provide an efficient method (although the implementation could be problematic). But, in practice, the complete specification of the model is unrealistic or impossible. First, the correlation structure over time needs to be specified. A particular correlation or time series structure tends to produce a heavily model-dependent method, and is difficult to handle, when the number of latent variables change over occasions, when the occasions are unequally spaced, or when the number of occasions is small. On the other hand, assuming the unrestricted general covariance structure for the repeated measures of the multivariate latent variables requires estimation of a large covariance matrix which may not be positive definite (e.g., some latent variables may not change over time), and can lead to identification and model fitting difficulties. Another reason making the specification of the model unrealistic is that the multivariate longitudinal data often include non-normal (highly skewed or highly discrete) variables. In practice, the multivariate observations, as well as underlying latent vectors, consist of mixed types of variables, i.e., discrete, normal, and unspecified non-normal variables. Thus, it is desirable to develop a model fitting procedure that can be applied without restrictive specifications of the correlation structure and the distributional form, that can be implemented readily, and that produces valid statistical inferences without much difficulty and without additional assumptions. This paper proposes such a procedure, and discusses its use and justification. The procedure is especially useful in the model building and checking process where the inferences for certain model parameters and model fit have to be made quickly and easily. This paper concentrates on the use of the proposed procedure for the purpose of making inferences for model parameters and model fit. But, the procedure can be used as the first step for modeling the time series structure of latent variables, or as a method for selecting a model to be used in search of more complete specification of the data structure.

The procedure proposed and discussed in this paper is called the pseudo-independence (PI) method. The method is based on a reduced likelihood, but uses all available observations. The PI can be implemented with ease, when the full likelihood approach is infeasible. In fact, the existing computer packages can be utilized to carry out the PI analysis. Also, the PI inference procedures for many of the relevant parameters and for checking model fit are valid in large samples for a broad class of non-normal data and for any type of individual trend over time. It is also shown that the efficiency loss of the PI method is minimal relative to the full likelihood method even when the latter can be implemented.

In Section 4.3, the PI method is introduced and the limiting distributions of the proposed estimator and goodness-of-fit statistic are presented under the assumption of normally distributed observations. The property and usefulness of the PI method under a general non-normal assumptions is discussed in Section 4.4. The efficiency of the method is considered and illustrated by a numerical example in Section 4.5. The proofs of all theorems are given in the Appendix.

4.3 Pseudo-Independence Method

Suppose that a population is monitored at T different occasions. At the t-th occasion ($t = 1, 2, \cdots, T$), $p^{(t)}$ measurements are taken from each of the $n^{(t)}$ individuals. Let $\mathbf{z}_i^{(t)}$ be a $p^{(t)} \times 1$ observation vector that contains the $p^{(t)}$ measurements from the i-th individual at the t-th occasion. It is assumed that the relationships among the observed variables (measurements) at the t-th occasion can be explained by $k^{(t)}$ factors. Let $\mathbf{f}_i^{(t)}$ denote a $k^{(t)} \times 1$ latent vector that contains the $k^{(t)}$ factors for the i-th individual at the t-th occasion. The factor analysis model for the t-th occasion is

$$\mathbf{z}_i^{(t)} = \boldsymbol{\beta}_0^{(t)} + \mathbf{B}^{(t)}\mathbf{f}_i^{(t)} + \mathbf{e}_i^{(t)}, \quad i = 1, 2, \cdots, n^{(t)}. \tag{4.1}$$

The mean vector of the random error is $\mathrm{E}\{\mathbf{e}_i^{(t)}\} = 0$, and the covariance matrix of $\mathbf{e}_i^{(t)}$ is a diagonal matrix, $\mathrm{Var}\{\mathbf{e}_i^{(t)}\} = \mathbf{\Psi}^{(t)} = \mathrm{diag}(\psi_1^{(t)}, \cdots, \psi_{p^{(t)}}^{(t)})$. The latent factor vectors, $\mathbf{f}_i^{(t)}$, are assumed to be random with $\mathrm{E}\{\mathbf{f}_i^{(t)}\} = \boldsymbol{\mu}_{\mathbf{f}}^{(t)}$, and $\mathrm{Var}\{\mathbf{f}_i^{(t)}\} = \mathbf{\Phi}^{(tt)}$. It is assumed that the factors and the errors are independent, that $\mathbf{e}_i^{(t)}$ are independent for all i and t, and that $\mathbf{f}_i^{(t)}$ are generally correlated over t.

In model (4.1) the dimension $p^{(t)}$ of the observation $\mathbf{z}_i^{(t)}$ can change over time t.

This allows situations where some measurements are not taken at some time points or the use of different instruments at different occasions. Accordingly, the dimension $k^{(t)}$ of the factor $\mathbf{f}_i^{(t)}$ can also vary over time. Thus, even the conceptual definition of the elements of $\mathbf{f}_i^{(t)}$ is allowed to change over t. Although the index for individuals is denoted by $i = 1, 2, \cdots, n^{(t)}$ instead of i_t for simplicity, the results would apply to any general unbalanced structure. For this general model, it is difficult to specify a particular dependency or correlations among $\mathbf{f}_i^{(t)}$ over t. Model (4.1) allows any general, unspecified dependency structure for $\mathbf{f}_i^{(t)}$. For the identification of (4.1) for $t = 1, 2, \cdots, T$, the following parameterization is conceptually and practically useful. Since individual traits or characteristics are expected to change over time, the factor mean $\boldsymbol{\mu}_{\mathbf{f}}^{(t)}$ and the covariance matrix $\boldsymbol{\Phi}^{(tt)}$ are generally different over time, and are treated as unrestricted. The factor vectors $\mathbf{f}_i^{(t)}$ for different individuals are assumed independent for all occasions, but $\mathbf{f}_i^{(t)}$ and $\mathbf{f}_i^{(m)}$ for an individual i at different occasions are generally correlated. Thus, the identification should be achieved through restrictions on $\boldsymbol{\beta}_0^{(t)}$, $\mathbf{B}^{(t)}$, and possibly $\boldsymbol{\Psi}^{(t)}$. To express a general form of such restrictions, we assume that $\boldsymbol{\beta}_0^{(t)}$, $\mathbf{B}^{(t)}$, and $\boldsymbol{\Psi}^{(t)}$ are functions of a vector $\boldsymbol{\tau}$ of dimension $d_{\boldsymbol{\tau}} \times 1$,

$$\boldsymbol{\beta}_0^{(t)} = \boldsymbol{\beta}_0^{(t)}(\boldsymbol{\tau}), \quad \mathbf{B}^{(t)} = \mathbf{B}^{(t)}(\boldsymbol{\tau}), \quad \boldsymbol{\Psi}^{(t)} = \boldsymbol{\Psi}^{(t)}(\boldsymbol{\tau}). \tag{4.2}$$

Let $\boldsymbol{\theta} = (\boldsymbol{\tau}', \boldsymbol{\mu}_f', \boldsymbol{\varphi}')'$, where $\boldsymbol{\mu}_f = (\boldsymbol{\mu}_f^{(1)\prime}, \cdots, \boldsymbol{\mu}_f^{(T)\prime})'$, $\boldsymbol{\varphi} = (\text{vech}'\boldsymbol{\Phi}^{(11)}, \cdots, \text{vech}'\boldsymbol{\Phi}^{(TT)})'$, and the vech notation is defined as follows. For a $p \times q$ matrix $\mathbf{A} = (\mathbf{a}_1, \cdots, \mathbf{a}_q)$, let $\text{vec}\mathbf{A} = (\mathbf{a}_1', \cdots, \mathbf{a}_q')'$ be the $pq \times 1$ vector which lists all elements of $\mathbf{A}$. For a $p \times p$ symmetric matrix $\mathbf{A}$, let $\text{vech}\mathbf{A}$ denote the $[p(p+1)/2] \times 1$ vector listing the elements of $\mathbf{A}$ on and below the diagonal starting with the first column. Note that $\boldsymbol{\theta}$ does not include the covariances of the factors between the occasions and other possible parameters, but that $\boldsymbol{\theta}$ contains all the parameters in

$$\begin{aligned} \mathrm{E}\{\mathbf{z}_i^{(t)}\} &= \boldsymbol{\mu}^{(t)}(\boldsymbol{\theta}) = \boldsymbol{\beta}_0^{(t)}(\boldsymbol{\tau}) + \mathbf{B}^{(t)}(\boldsymbol{\tau})\boldsymbol{\mu}_f^{(t)}, \\ \mathrm{Var}\{\mathbf{z}_i^{(t)}\} &= \boldsymbol{\Sigma}^{(t)}(\boldsymbol{\theta}) = \mathbf{B}^{(t)}(\boldsymbol{\tau})\boldsymbol{\Phi}^{(tt)}\mathbf{B}^{(t)\prime}(\boldsymbol{\tau}) + \boldsymbol{\Psi}^{(t)}(\boldsymbol{\tau}). \end{aligned} \tag{4.3}$$

We will propose an estimation procedure for $\boldsymbol{\theta}$, with emphasis on making simple but correct inferences on $\boldsymbol{\tau}$. The parameters $\boldsymbol{\beta}_0^{(t)}(\boldsymbol{\tau})$ and $\mathbf{B}^{(t)}(\boldsymbol{\tau})$ represent the relationships between the observed measurements and the underlying factors. Such relationships are associated with particular instruments or measurement procedures, and may stay unchanged or have a particular structure over time. On the other hand, $\boldsymbol{\Psi}^{(t)}$ is a covariance matrix of measurement errors $\mathbf{e}_i^{(t)}$, and some components of $\boldsymbol{\Psi}^{(t)}$ may stay constant over time if some measurement procedures are repeatedly applied. Thus, the parameter $\boldsymbol{\tau}$ is most relevant for modeling and for assessing measurement structure and relationships. Also the inference for $\boldsymbol{\tau}$ may be extended to a general measurement model beyond a given sample of individuals. It is useful and convenient, especially at a model-building stage of the analysis, to be able to make inferences for $\boldsymbol{\tau}$ without specifying the dependency structure for the factors $\mathbf{f}_i^{(t)}$ in a given sample. It will be shown that this type of parameterization provides certain desirable properties of statistical inference procedures in addition to allowing practically meaningful interpretation.

Let $\bar{\mathbf{z}}^{(t)}$ and $\mathbf{S}^{(t)}$ denote the sample mean vector and covariance matrix for the $t-$th occasion,

$$\bar{\mathbf{z}}^{(t)} = \frac{1}{n^{(t)}}\sum_{i=1}^{n^{(t)}}\mathbf{z}_i^{(t)}, \quad \mathbf{S}^{(t)} = \frac{1}{n^{(t)}-1}\sum_{i=1}^{n^{(t)}}(\mathbf{z}_i^{(t)} - \bar{\mathbf{z}}^{(t)})(\mathbf{z}_i^{(t)} - \bar{\mathbf{z}}^{(t)})'.$$

To estimate $\boldsymbol{\theta}$, consider $\hat{\boldsymbol{\theta}}$ obtained by minimizing the following function over the parameter space:

$$\begin{aligned} Q(\boldsymbol{\theta}) &= \sum_{t=1}^{T} n^{(t)}\{\mathrm{tr}\{\mathbf{S}^{(t)}\boldsymbol{\Sigma}^{(t)-1}(\boldsymbol{\theta})\} - \log|\mathbf{S}^{(t)}\boldsymbol{\Sigma}^{(t)-1}(\boldsymbol{\theta})| \\ &\quad -p^{(t)} + [\bar{\mathbf{z}}^{(t)} - \boldsymbol{\mu}^{(t)}(\boldsymbol{\theta})]'\boldsymbol{\Sigma}^{(t)-1}(\boldsymbol{\theta})[\bar{\mathbf{z}}^{(t)} - \boldsymbol{\mu}^{(t)}(\boldsymbol{\theta})]\}. \end{aligned} \tag{4.4}$$

Note that $Q(\boldsymbol{\theta})$ would be a version of the correct likelihood, if $\mathbf{f}_i^{(t)}$ were independent over time, and if all variables are normally distributed. For model (4.1) with general dependency of $\mathbf{f}_i^{(t)}$ over t, with general unbalanced configuration, and with possibly different dimensions of $\mathbf{f}_i^{(t)}$ and $\mathbf{z}_i^{(t)}$ over time, the full likelihood function based on the joint density of all observations does not have a simple form, even if all variables are assumed normal. For normal data, $Q(\boldsymbol{\theta})$ is -2 times the product of the marginal distributions of $\mathbf{z}_i^{(t)}$ from the $t-$th occasion, and is different from the full likelihood unless observations over time are independent. Thus, the use of $Q(\boldsymbol{\theta})$ for the longitudinal data corresponds to the case of an incorrect or reduced likelihood that contains relevant parameters specified in the model. We call this estimation procedure the pseudo-independence (PI) method, and $\hat{\boldsymbol{\theta}}$ the PI estimator of $\boldsymbol{\theta}$, since the method incorrectly assumes independence over time for convenience. The minimization of $Q(\boldsymbol{\theta})$ can be carried out easily using the multi-sample option in the existing computer packages such as EQS and LISREL that were developed for independent normal samples from multiple populations. Such programs also compute $Q(\hat{\boldsymbol{\theta}})$ as the goodness-of-fit test statistic. To assess the usefulness of the PI method in practice, we study the correctness of inference procedures based on the pseudo-independence and pseudo-normality as applied to longitudinally dependent non-normal data. In this section, we concentrate on normal data, and see how the standard errors of $\hat{\boldsymbol{\theta}}$ and the goodness-of-fit test statistic $Q(\hat{\boldsymbol{\theta}})$ are affected by incorrectly assuming the independence of observations over time.

To present the properties of $\hat{\boldsymbol{\theta}}$ and $Q(\hat{\boldsymbol{\theta}})$, we introduce some notation. For a $p \times p$ symmetric $\mathbf{A}$, define $\mathbf{D}_p$ to be the $p^2 \times [p(p+1)/2]$ matrix such that $\text{vec}\mathbf{A} = \mathbf{D}_p\text{vech}\mathbf{A}$, and $\mathbf{D}_p^+ = (\mathbf{D}_p'\mathbf{D}_p)^{-1}\mathbf{D}_p'$ so that $\text{vech}\mathbf{A} = \mathbf{D}_p^+\text{vec}\mathbf{A}$. See, e.g., Fuller (1987) and Magnus and Neudecker (1988). Also, let $\boldsymbol{\theta}_0$ is the true value of $\boldsymbol{\theta}$ corresponding to the true distribution generating data. Define

$$\mathbf{F}_0 = \mathbf{F}(\boldsymbol{\theta}_0) = [\mathbf{F}^{(1)\prime}(\boldsymbol{\theta}), \cdots, \mathbf{F}^{(T)\prime}(\boldsymbol{\theta})]'|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}, \quad \mathbf{F}^{(t)}(\boldsymbol{\theta}) = \begin{bmatrix} \frac{\partial \boldsymbol{\mu}^{(t)}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} \\ \frac{\partial \text{vech}\boldsymbol{\Sigma}^{(t)}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} \end{bmatrix}. \qquad (4.5)$$

Let

$$n = n^{(1)} + \cdots + n^{(T)}, \quad n_m = \min\{n^{(1)}, \cdots, n^{(T)}\},$$

and let $n^{(tm)}$ be the number of individuals measured at both t−th and m−th occasions. In deriving large sample results, we consider the limit as $n_m \to \infty$. For the general unbalanced model (4.1), specifying assumptions on the factors $\mathbf{f}_i^{(t)}$ requires some care. It is convenient for specifying assumptions to have a notation for all underlying factors at all occasions for an individual including factors at the time points where this individual is not measured. For this, for an individual i with at least one observation in the sample, let $\mathbf{f}_i = (\mathbf{f}_i^{(1)\prime}, \cdots, \mathbf{f}_i^{(T)\prime})'$, where some $\mathbf{f}_i^{(t)}$ may not exist in a given data set (i.e., observation $\mathbf{z}_i^{(t)}$ is missing). Write $\text{E}\{\mathbf{f}_i\} = \boldsymbol{\mu}_\mathbf{f} = (\boldsymbol{\mu}_\mathbf{f}^{(1)\prime}, \cdots, \boldsymbol{\mu}_\mathbf{f}^{(T)\prime})'$ and

$$\text{Var}\{\mathbf{f}_i\} = \boldsymbol{\Phi} = \begin{pmatrix} \boldsymbol{\Phi}^{(11)} & \ldots & \boldsymbol{\Phi}^{(1T)} \\ \vdots & \ddots & \vdots \\ \boldsymbol{\Phi}^{(T1)} & \ldots & \boldsymbol{\Phi}^{(TT)} \end{pmatrix}.$$

Throughout our development, the following regularity conditions for identification and asymptotic set-up are assumed for model (4.1)-(4.2):

Assumption 1 *i) For any $\varepsilon > 0$ there exists a $\delta > 0$ such that for $\boldsymbol{\theta}$ in the parameter space if $\sum_{t=1}^{T}[\| \boldsymbol{\mu}^{(t)}(\boldsymbol{\theta}) - \boldsymbol{\mu}^{(t)}(\boldsymbol{\theta}_0) \| + \| \boldsymbol{\Sigma}^{(t)}(\boldsymbol{\theta}) - \boldsymbol{\Sigma}^{(t)}(\boldsymbol{\theta}_0) \|] < \delta$ then $\| \boldsymbol{\theta} - \boldsymbol{\theta}_0 \| < \varepsilon$, where $\boldsymbol{\mu}^{(t)}(\boldsymbol{\theta})$ and $\boldsymbol{\Sigma}^{(t)}(\boldsymbol{\theta})$ are defined in (4.3), and $\| \mathbf{A} \| = \sqrt{tr\{\mathbf{A}'\mathbf{A}\}}$. Also, $\mathbf{F}_0$ defined in (4.5) has full column rank.*

ii) The $\boldsymbol{\beta}_0^{(t)}(\boldsymbol{\tau})$, $\mathbf{B}^{(t)}(\boldsymbol{\tau})$, and $\boldsymbol{\Psi}^{(t)}(\boldsymbol{\tau})$, for $t = 1, 2, \cdots, T$, are twice continuously differentiable in a neighborhood of $\boldsymbol{\tau}_0$, the true value of $\boldsymbol{\tau}$.

iii) The parameter space for $\boldsymbol{\tau}$ is an open set, the parameter space for $\boldsymbol{\mu}_{\mathbf{f}}$ is the set of all $(\sum_{t=1}^{T} k^{(t)}) \times 1$ vectors, and the true values $\boldsymbol{\Phi}_0^{(tt)}$, $t = 1, 2, \cdots, T$, are positive definite.

iv) $\lim_{n_m \to \infty}(n^{(t)}/n) = r^{(t)} > 0$, $t = 1, 2, \cdots, T$ and $\lim_{n_m \to \infty}[nn^{(tm)}/(n^{(t)}n^{(m)})] = \lambda^{(tm)}$, $t \neq m = 1, 2, \cdots, T$.

Assumption 1 *i*) is the standard identification assumption which is needed for our very general model (4.1) and (4.2). Assumption 1 *iii*) quarantees that the true value $\boldsymbol{\theta}_0$ is in the interior of the parameter space so that the limiting normal distribution of $\hat{\boldsymbol{\theta}}$ can be obtained. Assumption 1 *iv*) assumes that the sample sizes $n^{(t)}$, $t = 1, 2, \cdots, T$, increase at a common rate as $n_m \to \infty$.

For $\hat{\boldsymbol{\theta}}$ and $Q(\hat{\boldsymbol{\theta}})$ computed under the incorrect independence over time, the following theorem gives the limiting distributions when the general longitudinal dependency exists and all the variables follow normal distribution. The proof of this and other theorems are given in the appendix. The limiting covariance matrix $\mathbf{V}$ of $\hat{\boldsymbol{\theta}}$ is shown to be the sum of two matrices $\mathbf{V}_I$ and $\mathbf{V}_D$, where $\mathbf{V}_I$ would be the limiting covariance matrix if the factors were independent over time, and $\mathbf{V}_D$ is a matrix with zero diagonal elements and some zero sub-matrices.

Theorem 1 *Let model (4.1)-(4.2) and Assumption 1 hold. In addition, assume*

Assumption 2 $\mathbf{f}_i \sim \mathrm{N}(\boldsymbol{\mu}_{\mathbf{f}}, \boldsymbol{\Phi})$, $\mathbf{e}_i^{(t)} \sim \mathrm{N}(\mathbf{0}, \boldsymbol{\Psi}^{(t)})$, *and* $\mathbf{f}_i$*'s and* $\mathbf{e}_i^{(t)}$*'s are independent for all* i *and* t.

Then, the following holds.

a) As $n_m \to \infty$,

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{L} \mathrm{N}(\mathbf{0}, \mathbf{V}),$$

$$\mathbf{V} = \mathbf{V}_I + \mathbf{V}_D,$$

$$\mathbf{V}_I = (\mathbf{F}_0' \boldsymbol{\Omega}_0^{-1} \mathbf{F}_0)^{-1}, \quad \mathbf{V}_D = \begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_0 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{B}_0 \end{pmatrix},$$

where $\mathbf{F}_0$ *is defined in (4.5),*

$$\boldsymbol{\Omega}_0^{-1} = \boldsymbol{\Omega}^{-1}(\boldsymbol{\theta}_0) = blockdiag(r^{(1)}\boldsymbol{\Omega}^{(1)-1}(\boldsymbol{\theta}_0), \cdots, r^{(T)}\boldsymbol{\Omega}^{(T)-1}(\boldsymbol{\theta}_0)),$$

$$\boldsymbol{\Omega}^{(t)-1}(\boldsymbol{\theta}) = \begin{pmatrix} \boldsymbol{\Sigma}^{(t)-1}(\boldsymbol{\theta}) & \mathbf{0} \\ \mathbf{0} & \mathbf{G}^{(t)-1}(\boldsymbol{\theta}) \end{pmatrix},$$

$$\mathbf{G}^{(t)-1}(\boldsymbol{\theta}) = \frac{1}{2}\mathbf{D}_{p^{(t)}}'[\boldsymbol{\Sigma}^{(t)-1}(\boldsymbol{\theta}) \otimes \boldsymbol{\Sigma}^{(t)-1}(\boldsymbol{\theta})]\mathbf{D}_{p^{(t)}},$$

$\mathbf{A}_0 = [\mathbf{A}_0^{(tm)}]$ *and* $\mathbf{B}_0 = [\mathbf{B}_0^{(tm)}]$ *are partitioned matrices with* $k^{(t)} \times k^{(m)}$ $\mathbf{A}_0^{(tm)}$, *and* $k_*^{(t)} \times k_*^{(m)}$ $\mathbf{B}_0^{(tm)}$ ($k_*^{(t)} = k^{(t)}(k^{(t)}+1)/2$), *and satisfy*

$$\mathbf{A}_0^{(tm)} = \begin{cases} \mathbf{0}, & \text{if } t = m, \\ \lambda^{(tm)}\boldsymbol{\Phi}_0^{(tm)}, & \text{if } t \neq m, \end{cases}$$

$$\mathbf{B}_0^{(tm)} = \begin{cases} \mathbf{0}, & \text{if } t = m, \\ \lambda^{(tm)}\mathbf{D}_{p^{(t)}}^{+}[\boldsymbol{\Phi}_0^{(tm)} \otimes \boldsymbol{\Phi}_0^{(tm)}]\mathbf{D}_{p^{(m)}}^{+\prime}, & \text{if } t \neq m, \end{cases}$$

and $\Phi_0^{(tm)}$, $t \neq m = 1, 2, \cdots, T$ *is the true value of the covariance matrix between* $\mathbf{f}_i^{(t)}$ *and* $\mathbf{f}_i^{(m)}$.

b) As $n_m \to \infty$, $Q(\hat{\boldsymbol{\theta}}) \xrightarrow{L} \chi_q^2$ *where* $Q(\boldsymbol{\theta})$ *is defined in (4.4), and*

$$q = \sum_{t=1}^{T}[p^{(t)} + \frac{p^{(t)}(p^{(t)}+1)}{2}] - d_{\boldsymbol{\tau}} - \sum_{t=1}^{T}[k^{(t)} + \frac{k^{(t)}(k^{(t)}+1)}{2}].$$

By Theorem 1(a), the diagonal elements of $\mathbf{V}$ and $\mathbf{V}_I$ are the same, because $\mathbf{V}_D$ has 0 diagonal elements. Hence, the asymptotic standard errors for all components of $\hat{\boldsymbol{\theta}}$ computed under the incorrect assumption of independence over time are in fact correct for the longitudinal data with dependency over time. Since the part of $\mathbf{V}_D$ corresponding to $\hat{\boldsymbol{\tau}}$ is $\mathbf{0}$, the asymptotic inferences concerning any functions of $\boldsymbol{\tau}$ can be made based on $\hat{\boldsymbol{\tau}}$ and $\mathbf{V}_I$. Recall that $\boldsymbol{\tau}$ contains all coefficient/relationship parameters and the error variances, both possibly restricted over time. For a given t, the asymptotic inferences for any functions of $(\boldsymbol{\mu}_{\mathbf{f}}^{(t)}, \boldsymbol{\Phi}^{(tt)})$ can still be made correctly using $\mathbf{V}_I$, because $\mathbf{V}_D$ is block diagonal. The estimated asymptotic covariance matrix of the PI estimator $\hat{\boldsymbol{\theta}}$ is computed based on $\mathbf{V}_I$ by the existing packages for the independent multi-sample analysis. Therefore, the use of the PI estimator $\hat{\boldsymbol{\theta}}$ and the pseudo-independence covariance matrix $\mathbf{V}_I$ for the longitudinal data can be carried out easily by the existing packages, and is asymptotically valid for most types of inferences, if all variables are normally distributed, and if the parameterization of model (4.1)-(4.2) is used. This result is extended to non-normal data in the next section.

4.4 Non-normal Data

In Section 4.3, it was assumed that all the variables follow normal distribution. In this section, the assumption of normality is relaxed, and we show that parts of Theorem 1 hold under more general assumptions and conditions. In particular, we show that some asymptotic inference procedures derived under the pseudo normal-independence model are valid for a broad range of non-normal, longitudinally dependent data. In the literature of factor analysis and latent variable modeling, the property that some estimators and test statistics have a common limiting distribution under different sets of assumptions is refered to as asymptotic robustness.

Amemiya, Fuller and Pantula (1987) considered exploratory factor analysis and proved that the asymptotic standard errors of some estimators are the same for fixed, non-normal, and normal factors assuming normality for the errors. Anderson and Amemiya (1988) extended the above results to confirmatory factor analysis and non-normal errors. Amemiya and Anderson (1990) showed that some goodness-of-fit statistics have a common limiting distribution for non-normal factors and errors. Browne and Shapiro (1988) and Anderson (1987, 1989) extended the above results for more general models that include structural equation modeling. Browne (1990) and Satorra (1992) considered models including mean and covariance structures. Asymptotic robustness for the multi-sample analysis was considered by Satorra (1993,1994) and Papadopoulos and Amemiya (1994, 1996). The asymptotic robustness for longitudinal factor analysis has not been discussed in the literature. The next theorem shows that the limiting distributions of $\hat{\tau}$ and $Q(\hat{\theta})$ derived in Theorem 1 under the assumption of normality for $\mathbf{f}_i^{(t)}$ and $\mathbf{e}_i^{(t)}$ are the same for any non-normal $\mathbf{f}_i^{(t)}$ with finite second moments.

Theorem 2 *Let model (4.1)-(4.2) and Assumption 1 hold. In addition, assume* **Assumption 2-a)** $\mathbf{f}_i$*'s are independently and identically distributed (i.i.d.) with* $E\{\mathbf{f}_i\} = \boldsymbol{\mu}_\mathbf{f}$ *and* $Var\{\mathbf{f}_i\} = \boldsymbol{\Phi}$, $\mathbf{e}_i^{(t)} \sim \mathbf{N}(\mathbf{0}, \boldsymbol{\Psi}^{(t)})$, *and* $\mathbf{f}_i$*'s and* $\mathbf{e}_i^{(t)}$*'s are independent for all* i *and* t.

Then, as $n_m \to \infty$,

a)

$$\sqrt{n}(\hat{\boldsymbol{\tau}} - \boldsymbol{\tau}_0) \xrightarrow{L} \mathbf{N}(\mathbf{0}, \mathbf{V}_I^{\boldsymbol{\tau}}),$$

where $\mathbf{V}_I^{\boldsymbol{\tau}}$ *is the part of* $\mathbf{V}_I$ *corresponding to* $\boldsymbol{\tau}$,

b) $Q(\hat{\boldsymbol{\theta}}) \xrightarrow{L} \chi_q^2$ *where* q *is given in Theorem 1-b).*

In Assumption 2-a), the errors $\mathbf{e}_i^{(t)}$ are normally distributed, but $\mathbf{f}_i^{(t)}$ have unspecified distribution and longitudinal dependancy. Theorem 2-a) states that the asymptotic inferences fot $\boldsymbol{\tau}$ based on $\hat{\boldsymbol{\tau}}$ and $\mathbf{V}_I^{\boldsymbol{\tau}}$ for the psuedo normal-independence model are valid for any non-normal factors $\mathbf{f}_i^{(t)}$. Theorem 2-b) implies that the goodness-of-fit test using $Q(\hat{\boldsymbol{\theta}})$ and χ_q^2 distribution can be used correctly to test the fit of model (4.1) without specifying the distribution of longitudinally dependent $\mathbf{f}_i^{(t)}$.

A result similar to Theorem 2 can be obtained for non-normal $\mathbf{e}_i^{(t)}$ (in addition to non-normal $\mathbf{f}_i^{(t)}$), provided that the $p^{(t)}$ elements $e_{ik}^{(t)}$, $k = 1, 2, \cdots, p^{(t)}$, of $\mathbf{e}_i^{(t)}$ are independent and have unrestricted variances $\psi_k^{(t)}$ ($\sum_{t=1}^{T} p^{(t)}$ error variances are unrestricted and to be estimated). Under such a condition, we write

$$\boldsymbol{\tau} = (\boldsymbol{\tau}_1', \ \boldsymbol{\psi}')', \quad \boldsymbol{\beta}_0^{(t)} = \boldsymbol{\beta}_0^{(t)}(\boldsymbol{\tau}_1), \quad \mathbf{B}^{(t)} = \mathbf{B}^{(t)}(\boldsymbol{\tau}_1), \tag{4.6}$$

where $\boldsymbol{\psi}$ includes $\sum_{t=1}^{T} p^{(t)}$ error variances, i.e., $\boldsymbol{\psi} = (\boldsymbol{\psi}^{(1)\prime}, \cdots, \boldsymbol{\psi}^{(T)\prime})'$ and $\boldsymbol{\psi}^{(t)} = (\psi_1^{(t)}, \cdots, \psi_{p^{(t)}}^{(t)})'$, and no relation is assumed between $\boldsymbol{\tau}_1$ and $\boldsymbol{\psi}$. The result is derived in the next theorem. It is understood that Assumption 1 *ii)* and *iii)* apply to $\boldsymbol{\tau}$ of the form (4.6). Thus, $\boldsymbol{\beta}_0^{(t)}(\boldsymbol{\tau}_1)$ and $\mathbf{B}^{(t)}(\boldsymbol{\tau}_1)$ are twice continuously differentiable in a neighborhood of the true value $\boldsymbol{\tau}_{10}$, and the true error variances $\psi_{k0}^{(t)}$ are strictly positive.

Theorem 3 *Let model (4.1)-(4.2) hold with* $\boldsymbol{\beta}_0^{(t)} = \boldsymbol{\beta}_0^{(t)}(\boldsymbol{\tau}_1)$ *and* $\mathbf{B}^{(t)} = \mathbf{B}^{(t)}(\boldsymbol{\tau}_1)$. *Let Assumption 1 hold. In addition, assume*

Assumption 2-b) $\mathbf{f}_i$*'s are i.i.d. with* $E\{\mathbf{f}_i\} = \boldsymbol{\mu}_{\mathbf{f}}$ *and* $Var\{\mathbf{f}_i\} = \boldsymbol{\Phi}$, *and* $\mathbf{f}_i$*'s and* $e_{ik}^{(t)}$*'s are independent for aii i, t, and k. For a given t,* $\{\mathbf{e}_i^{(t)},\ i = 1, 2, \cdots, n^{(t)}\}$ *are i.i.d. with* $E\{e_{ik}^{(t)}\} = 0$ *and* $Var\{e_{ik}^{(t)}\} = \psi_k^{(t)}$.

Then, as $n_m \to \infty$,

a)

$$\sqrt{n}(\hat{\boldsymbol{\tau}}_1 - \boldsymbol{\tau}_{10}) \xrightarrow{L} \mathbf{N}(\mathbf{0}, \mathbf{V}_I^{\boldsymbol{\tau}_1}),$$

where $\mathbf{V}_I^{\boldsymbol{\tau}_1}$ *is the part of* $\mathbf{V}_I$ *corresponding to* $\boldsymbol{\tau}_1$,

b) $Q(\hat{\boldsymbol{\theta}}) \xrightarrow{L} \chi_q^2$ *where q is given in Theorem 1-b).*

Theorem 3 has shown that the asymptotic inferences for the relationship coefficient parameter $\boldsymbol{\tau}_1$ based on the pseudo normal-independence model $\hat{\boldsymbol{\tau}}_1$ and $\mathbf{V}_I^{\boldsymbol{\tau}_1}$ as well as the test for model fit based on $Q(\hat{\boldsymbol{\theta}})$ are valid for non-normal factors and non-normal errors, provided all error variances are unrestricted and estimated. Note that the limiting distribution of $\hat{\boldsymbol{\psi}}$ part of $\hat{\boldsymbol{\tau}}$ depends on the distribution of $e_{ik}^{(t)}$ and is not covered in Theorem 3.

One practical case not covered by Theorems 2 and 3 is that with non-normal errors $e_{ik}^{(t)}$ having equal variances over time. Suppose that some or all variables are measured repeatedly over time using the same instrument. Then, it is reasonable to restrict the error variances for such variables to be equal over time, e.g., $\psi_k^{(1)} = \psi_k^{(2)} = \cdots = \psi_k^{(T)}$ for the k-th variable. If the normality of the errors $e_{ik}^{(t)}$ can not be assumed, Theorem 2 does not apply. Theorem 3 is not applicable if $\psi_k^{(t)}$'s are restricted in the estimation procedure. For such a case with non-normal $e_{ik}^{(t)}$ with restricted variances, we can consider two approaches in making inferences about $\boldsymbol{\tau}_1$ defined in 4.6, both of which can be carried out using the pseudo normal-independence method.

In the first approach, the model with restricted $\psi_k^{(t)}$'s is fitted using the PI method, and asymptotic inferences for $\boldsymbol{\tau}_1$ are made based on $\mathbf{V}_I^{\boldsymbol{\tau}_1}$ in Theorem 3 despite the non-normality of $e_{ik}^{(t)}$'s. The inferences in this approach are not asymptotically correct, because Theorem 3 does not apply, and because the true limiting covariance matrix is not $\mathbf{V}_I^{\boldsymbol{\tau}_1}$. As illustrated later, the error made by this use of an incorrect asymptotic covariance matrix is negligible in most practical situations. The second approach uses the PI method to fit the model with completely unrestricted $\psi_k^{(t)}$'s despite the knowledge of restrictions among $\psi_k^{(t)}$'s. The second approach may not be efficient in the sense that unnecessarily too many parameters are being estimated. But, inferences for $\boldsymbol{\tau}_1$ based on $\mathbf{V}_I^{\boldsymbol{\tau}_1}$ in this approach are asymptotically valid for non-normal $e_{ik}^{(t)}$ because of Theorem 3. As the following illustration shows, the efficiency loss in estimating unrestricted $\psi_k^{(t)}$'s is minimal in most practical situations. In fact, depending on the kurtosis of non-normal $e_{ik}^{(t)}$, it is possible that the limiting variance of $\hat{\boldsymbol{\tau}}_1$ in the second approach is actually smaller than that in model fitting using the restrictions among $\psi_k^{(t)}$'s. This is because the efficiency of the pseudo normal-independence method fitting unrestricted non-normal $e_{ik}^{(t)}$ is the same as the case with normal $e_{ik}^{(t)}$, and because the efficiency fitting restricted non-normal $e_{ik}^{(t)}$ is lower than the normal case.

To explain intuitive reasons for the negligible error in the first approach and the negligible efficiency loss in the second, we note that the limiting covariance matrix of $\hat{\boldsymbol{\tau}}_1$ is generally the sum of two terms (following the argument of Theorem 3.R in Amemiya, Fuller, and Pantula (1987)). The first term is the limiting covariance matrix of $\hat{\boldsymbol{\tau}}_1$ when the error variances are known, and the second term represents the additional variability due to the estimation of unknown error variances. The first term stays the same for different non-normal distributions for $\mathbf{f}_i^{(t)}$ and $e_{ik}^{(t)}$ (asymptotic robustness), and is not affected by how the error variances are estimated. The second term is free of the distribution of $\mathbf{f}_i^{(t)}$, has different forms depending on different distributions of $e_{ik}^{(t)}$, and is larger if more unrestricted error variances are to be estimated. Hence, only the second term, not the first term, in the asymptotic covariance matrix of $\hat{\boldsymbol{\tau}}_1$ contributes to the error in using an incorrect asymptotic covariance matrix in the first approach, and to the asymptotic efficiency loss in estimating $\hat{\boldsymbol{\tau}}_1$ in the second approach. The second term depends on the kurtosis of the distribution of $e_{ik}^{(t)}$. Thus, the increase in the second term fitting too many $\psi_k^{(t)}$'s by the second approach can be smaller than the difference between the second terms for the cases with normal and non-normal $e_{ik}^{(t)}$. Since the second term is an additional variability term representing estimation error in $\hat{\boldsymbol{\psi}}$, it is generally smaller than the first term. This is why the inference error in the first approach and the efficiency loss in the second approach are generally small. The second term in the asymptotic covariance matrix of $\hat{\boldsymbol{\tau}}_1$ is especially small compared to the first term if the error variances $\psi_k^{(t)}$

are small relative to the variability of the factor part involving $\mathbf{f}_i^{(t)}$. Thus, if we define the reliability of an observed variable $z_{ik}^{(t)}$ to be

$$R_k^{(t)} = 1 - \frac{\psi_k^{(t)}}{\mathrm{Var}\{z_{ik}^{(t)}\}}, \tag{4.7}$$

then the asymptotic inference error in the first approach and the efficiency loss in the second approach are both negligible unless the reliabilities are very small. Therefore, the PI method using either approach can still provide a useful and simple inference procedures for $\boldsymbol{\tau}_1$ even when the errors are non-normal and the error variances are known to be restricted. These points are illustrated by the following example.

For two time points, $t = 1, 2$, and for a 4×1 observed vectors $\mathbf{z}_i^{(t)}$, consider a longitudinal factor analysis model given by

$$\mathbf{z}_i^{(t)} = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ 1 \end{pmatrix} f_i^{(t)} + \mathbf{e}_i^{(t)},$$

where $f_i^{(t)}$ is a scalar, $\mathrm{Var}\{f_i^{(t)}\} = \varphi^{(tt)}$, $\mathrm{Var}\{\mathbf{e}_i^{(t)}\} = \mathrm{diag}(\psi_1^{(t)}, \cdots, \psi_4^{(t)})$, and $f_i^{(t)}$ and $e_{ik}^{(t)}$ are all scaled χ_1^2 random variables adjusted to have zero mean and appropriate variances. It is assumed to be known that $\psi_k^{(1)} = \psi_k^{(2)}$, $k = 1, 2, 3, 4$. The true values of the parameters are $\beta_m = 1$, $m = 1, 2, 3$, $\varphi^{(11)} = 1$, $\varphi^{(22)} = 2$. The parameters $\psi_k^{(t)}$, $k = 1, 2, 3, 4$, have a common true value ψ_0. Then, the reliability of each element of $\mathbf{z}_i^{(t)}$ is

$$R^{(1)} = \frac{1}{1 + \psi_0}, \quad t = 1, \qquad R^{(2)} = \frac{2}{2 + \psi_0}, \quad t = 2.$$

We used five different values of ψ_0 corresponding to ($R^{(1)}$, $R^{(2)}$)=(0.5, 0.67), (0.6, 0.75), (0.7, 0.82), (0.8, 0.89), (0.9, 0.95).

For the first approach described above, the model with known restriction $\psi_k^{(1)} = \psi_k^{(2)}$, $k = 1, 2, 3, 4$, is fitted using the PI method, estimating β_m's, $\varphi^{(11)}$, $\varphi^{(22)}$, and four error variances common over time. We computed two limiting standard deviations of $\hat{\beta}_m$ (common for $m = 1, 2, 3$) evaluated at the true values of the parameters. One is the square root of $\mathbf{V}_I^{\mathcal{T}_1}$ in Theorem 3 derived under an incorrect normal assumption for $e_{ik}^{(t)}$, and the other is the correct limiting standard error obtained using the properties of χ_1^2 distributions for $e_{ik}^{(t)}$. The ratio of the normal case standard error over the correct one is given in Table 4.1 for the five sets of ($R^{(1)}$, $R^{(2)}$). All ratios are very close to 1, and the ratio approaches 1 as the reliability increases. Thus, the error in using an incorrect asymptotic standard error based on $\mathbf{V}_I^{\mathcal{T}_1}$ and the PI method is negligible for the reliability as low as 0.5.

Table 4.1: Ratio for the limiting standard deviation (l.s.d.) of $\hat{\beta}_m$ for the first approach
[Ratio=(l.s.d. for $\hat{\beta}_m$ assuming normality)/ (correct l.s.d. for $\hat{\beta}_m$), $R^{(t)}$=reliability]

$R^{(1)}$	0.5	0.6	0.7	0.8	0.9
$R^{(2)}$	0.67	0.75	0.82	0.89	0.95
Ratio	0.998554	0.999318	0.999842	0.999957	0.999995

For the second approach, the issue is how much efficiency is lost (if any) in the estimation of β_m by ignoring the known equality of $\psi_k^{(t)}$'s over t, and by estimating different error variances over t. Thus, the second approach using the PI method estimates β_m's, $\varphi^{(11)}$, $\varphi^{(22)}$, and eight error variances $\psi_k^{(t)}$. By Theorem 3, the limiting standard deviation for this $\hat{\beta}_m$ (common for $m = 1, 2, 3$) is the square root of the diagonal element of $\mathbf{V}_I^{\mathcal{T}_1}$ for the model with eight error variances to be estimated.

This standard deviation is compared to the limiting standard deviations of two other estimators of β_m obtained by fitting equal $\psi_k^{(t)}$ over t (only four error variances). For the data used here, the correct standard deviation for the PI method with equal $\psi_k^{(t)}$ uses the χ_1^2 distribution property. This is the standard deviation used in the denominator of the ratio used in Table 4.1. The ratio of the standard deviation for fitting equal $\psi_k^{(t)}$ over t based on χ_1^2 distribution over that for fitting unequal $\psi_k^{(t)}$ over time is reported as Ratio (χ_1^2) in Table 4.2. As seen in Table 4.2, this ratio is in fact larger than 1. Note that the kurtosis of χ_1^2 is larger than that of normal. Thus, the efficiency gained by fitting a smaller number of error variances is offset by the efficiency loss due to the use of the normality-based PI method for χ_1^2 variables. The PI method with unequal error variances does not lose efficiency for non-normal $e_{ik}^{(t)}$ in the sense that the limiting standard deviation is common for all distributions. Thus, depending on the kurtosis of the errors, the PI method fitting unnecessarily too many error variances can be more efficient than that fitting restricted error variances, in terms of asymptotic inferences for $\boldsymbol{\tau}_1$. The other standard deviation compared to that for the second approach is the numerator of the ratio in Table 4.1. This would be the correct and the most efficient limiting standard deviation if $e_{ik}^{(t)}$'s were normally distributed. Thus, Ratio (N) in Table 4.2 of the standard deviation using equal $\psi_k^{(t)}$ over t and normality to the second approach standard deviation is known to be less than 1, and represents the efficiency loss of the second approach compared to the most efficient method for the case with normal errors. The values of Ratio (N) are all nearly 1, indicating that the efficiency loss by estimating many error variances is negligible even for the normal-error case. Hence, the first or second approach, both using the PI method, can be used in practice without serious inference error or efficiency loss, provided the sample sizes are large and the reliabilities are not very small.

Table 4.2: Ratios for the limiting standard deviation (l.s.d.) of $\hat{\beta}_m$ for the second approach.
[Ratio (χ_1^2)=(correct l.s.d. for $\hat{\beta}_m$)/ (l.s.d. for $\hat{\beta}_m$ by fitting unequal $\psi_k^{(t)}$ over t),
Ratio (N)=(l.s.d. for $\hat{\beta}_m$ by fitting equal $\psi_k^{(t)}$ over t with normality)/(l.s.d. for $\hat{\beta}_m$ by fitting unequal $\psi_k^{(t)}$ over t),
$R^{(t)}$=reliability]

$R^{(1)}$	0.5	0.6	0.7	0.8	0.9
$R^{(2)}$	0.67	0.75	0.82	0.89	0.95
Ratio (χ_1^2)	1.000896	1.000468	1.000089	1.000026	1.000004
Ratio (N)	0.999449	0.999786	0.999931	0.999983	0.999999

4.5 Efficiency Comparison

In this section, the efficiency of the pseudo-independence (PI) method is discussed in comparison to the full likelihood approach. The PI method uses a reduced form of the likelihood function. We proposed this method especially for unbalanced data where the full likelihood may not be simple to handle. Intuitively, the PI method may produce estimators with larger asymptotic standard errors than the full likelihood approach. In other words, there may be some efficiency loss. In the case of balanced normal data, the efficiency loss can be evaluated, since the full likelihood method can be applied. The efficiency loss of the PI method relative to the full likelihood in this case depends on the reliability of the observed measurements, the size of the correlation over time, and the difference between the degrees of freedom of the full likelihood method and the PI method. The number of error degrees of freedom for the PI method is q given in Theorem 1 b). The number of error degrees of freedom for the full likelihood method is

$$\begin{aligned} q^* &= [p + \frac{p(p+1)}{2} - d_T - k - \frac{k(k+1)}{2}] \\ &= q + w, \end{aligned}$$

where $p = \sum_{t=1}^{T} p^{(t)}$, $k = \sum_{t=1}^{T} k^{(t)}$, and

$$w = \sum_{t<m} [p^{(t)} p^{(m)} - k^{(t)} k^{(m)}]. \tag{4.8}$$

The difference w is positive, depends on the numbers of measurements ($p^{(t)}$) and factors ($k^{(t)}$), and increases as the number of occasions increases. A general claim is that the efficiency loss is non-trivial only if the reliability is very small, the correlation of the factors over time is very high, and w is very large, and that, otherwise, the efficiency loss is negligible. Although formulating and verifying this claim mathematically is not simple, the following numerical example can serve as an illustration.

Consider the model

$$\mathbf{z}_i^{(t)} = \begin{pmatrix} \beta_1 \\ \beta_2 \\ 1 \end{pmatrix} f_i^{(t)} + \mathbf{e}_i^{(t)},$$

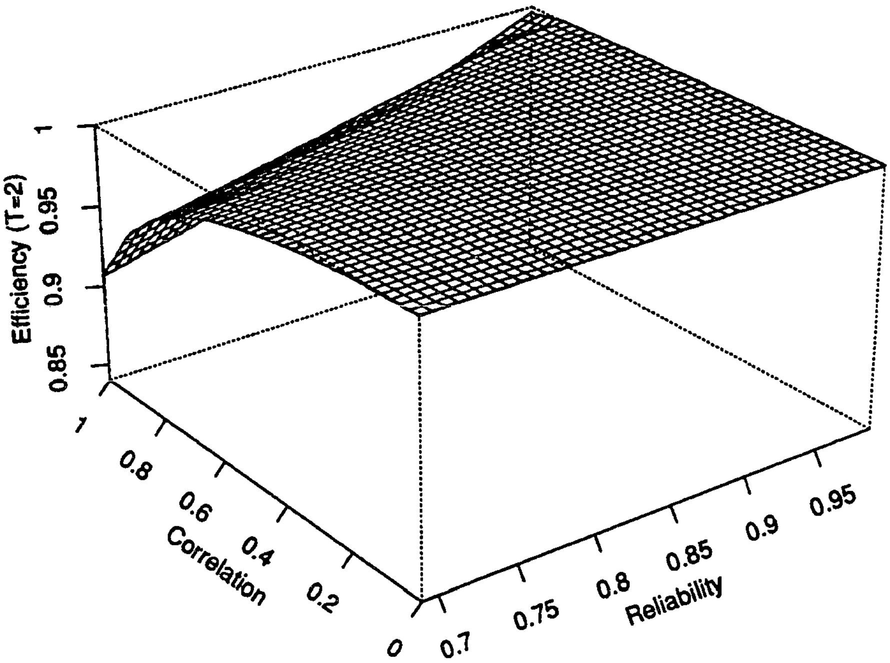
where the $\mathbf{z}_i^{(t)}$, $t = 1, 2, \cdots, T$, are 3×1 vectors with mean $\mathbf{0}$, $f_i^{(t)}$ are scalars with $\mathrm{Cov}\{f_i^{(t)}, f_i^{(m)}\} = \varphi_f^{(tm)}$ $t, m = 1, 2, \cdots, T$, and $\mathrm{Var}\{\mathbf{e}_i^{(t)}\} = \mathrm{diag}(\psi_1, \psi_2, \psi_3)$. We compute the limiting standard deviations for the PI and the full likelihood methods under the assumption that both $f_i^{(t)}$ and $\mathbf{e}_i^{(t)}$ are normal. The PI method estimates

$$\boldsymbol{\theta} = (\beta_1, \beta_2, \psi_1, \psi_2, \psi_3, \varphi_f^{(11)}, \cdots, \varphi_f^{(TT)}),$$

while in the full likelihood method the covariances of the factors, $\varphi_f^{(tm)}$ $t \neq m$, are estimated in addition to $\boldsymbol{\theta}$. The true values of β_1 and β_2 are 1 and -1 respectively, and the true $\varphi_f^{(tt)}$ $t = 1, 2, \cdots, T$ are all equal to 1. The true values of the variances of $e_{ik}^{(t)}$, $k = 1, 2, 3$, are all equal, and the common value is varied to produce the reliability (common for all $z_{ik}^{(t)}$, $k = 1, 2, 3$) ranging from 0.69 to 0.99 with an increment 0.03. The reliability is defined (4.7). The true values of the factor covariances $\varphi_f^{(tm)}$, $t \neq m$, are all set to ρ, and the value of ρ was varied from 0 to 1 with an increment of 0.1. All these cases were repeated for two and five occasions (T=2,5). In this example, w in (4.8) is equal to 8 for T=2, and equal to 80 for T=5. For estimating the parameter β_m, we computed the limiting standard deviations (evaluated at the true values) of the full likelihood and the PI estimators (common for β_1 and β_2). We define the efficiency of the PI method to be the ratio of the limiting standard deviation for the full likelihood method over that for the PI method. In Figure 4.1, the efficiency is plotted against the correlation and the reliability, for T=2 and T=5. Table 4.3 gives the efficiency for some cases, where the correlation (ρ) is equal to 0.6 and 0.8, and the reliability (R) is equal to 0.75 and 0.95. In Table 4.3, all the numbers are very close to 1, implying that the limiting standard deviation of the PI estimator are only slightly larger than those of the full likelihood estimator. It is clear from Figure 4.1 that the efficiency loss due to the use of the PI method is minimal for most practical situations even when the full likelihood method is possible. The simpler applicability for unbalanced cases and the asymptotic robustness properties make the PI method useful for factor analysis of longitudinal data in general.

Table 4.3: Efficiency of the PI method in comparison to the full likelihood method. (ρ=factor correlation over occasions, R=reliability, T=number of occasions).

	T=2		T=5	
	R=0.75	R=0.95	R=0.75	R=0.95
ρ=0.6	0.9892	0.9995	0.9804	0.9991
ρ=0.8	0.9712	0.9973	0.9646	0.9934



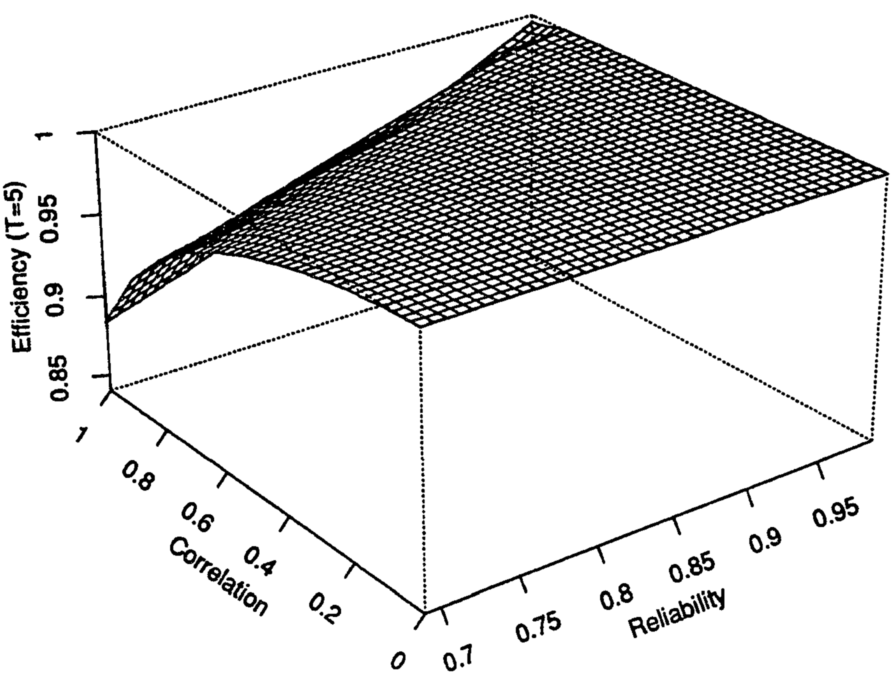


Figure 4.1: Efficiency of the PI Method in Comparison to the Full-Likelihood Method. (factor correlation over occasions, measurement reliability, T =number of occasions)

4.6 References

Allison, P. D. (1987). Estimation of linear models with incomplete data. *Sociological Methodology.* Washington, D.C.: American Sociological Association, pp. 71-103.

Amemiya, Y., Fuller W.A. and Pantula, S.G. (1987). The asymptotic distributions of some estimators for a factor analysis model. *Journal of Multivariate Analysis*, **22**, 51-64.

Amemiya, Y. and Anderson, T.W. (1990). Asymptotic chi-square tests for a large class of factor analysis models. *The Annals of Statistics*, **18**, 1453-1463.

Anderson, T.W. (1987). Multivariate linear relations. *Proceedings of the Second International Conference in Statistics,* edited by Pukkila T. and Puntanen S., University of Tampere, Finland. 9-36.

Anderson, T.W. (1989). Linear latent variable models and covariance structures. *Journal of Econometrics*, **41**, 91-119.

Anderson, T.W. and Amemiya, Y. (1988). The asymptotic normal distribution of estimators in factor analysis under general conditions. *The Annals of Statistics*, **16**, 759-771.

Baker, L.A., and Fulker, D.W. (1983). Incomplete covariance matrices and LISREL. *Data Analyst*, **1**, 3-5.

Basilevsky, A. (1994). *Statistical factor analysis and related methods.* John Wiley and Sons, New York.

Bentler, P.M. (1989). *EQS Structural Equations Program Manual.* BMDP Statistical Software, Inc., Los Angeles.

Bollen, P.A. (1989). *Structural Equations with Latent Variables.* John Wiley and Sons, New York.

Browne, M.W. (1990). Asymptotic robustness of normal theory methods for the analysis of latent curves. *Contemporary Mathematics*, **112**, 211-225.

Browne, M.W. and Shapiro, A. (1988). Robustness of normal theory methods in the analysis of linear latent variate models. *British Journal of Mathematical and Statistical Society*, **41**, 193-208.

Fuller, W.A. (1987). *Measurement Error Models*. John Wiley and Sons, New York.

Jöreskog, K. and Sörbom, D. (1989). *LISREL 7; A Guide to the Program and Applications*. 2nd ed., SPSS INC., Chicago.

Magnus, J. and Neudecker, H. (1988). *Matrix Differential Calculus*. John Wiley and Sons, New York.

Molenaar, C.M.P. (1985). A dynamic factor model for the analysis of multivariate time series. *Psychometrika*, **50**, 181-202.

Molenaar, C.M.P. (1992). Dynamic factor analysis model of non-stationary multivariate time series. *Psychometrika*, **57**, 333-349.

Papadopoulos, S. and Amemiya, Y. (1994). Asymptotic robustness for the structural equation analysis of several populations. *ASA Proceedings*, Business and Economics Statistics Section, 65-70.

Papadopoulos, S. and Amemiya, Y. (1996). On linear latent variable analysis of multiple populations. Unpublished. Iowa State University, Ames, Iowa.

Satorra, A. (1992). Asymptotic robust inferences in the analysis of mean and covariance structures. *Sociological Methodology,* edited by Marsden, P.V., 249-278.

Satorra, A. (1993). Asymptotic robust inferences multi-sample analysis of augmented-moment structures. *Multivariate Analysis: Future Directions 2*, edited by Cuadras, C.M. and Rao, C.R., 211-229.

Satorra, A. (1994). On asymptotic robustness in multiple-group analysis of multivariate relations. Paper presented at *Latent Variable Modeling with Applications to Causality*, March 19-20, Los Angeles.

Werts, C.E., Rock, D.A., and Grandy, J. (1979). Confirmatory factor analysis applications: Missing data problems and comparisons of path models between populations. *Multivariate Behavioral Research*, **14**, 199-213.

4.7 Appendix

Proof of Theorem 1

a) Under Assumption 2, $\bar{\mathbf{z}}^{(t)} \xrightarrow{p} \boldsymbol{\mu}^{(t)}(\boldsymbol{\theta}_0)$, $\mathbf{S}^{(t)} \xrightarrow{p} \boldsymbol{\Sigma}^{(t)}(\boldsymbol{\theta}_0)$. Thus, $Q(\boldsymbol{\theta})$ in (4.4) satisfies $Q(\boldsymbol{\theta}_0) \xrightarrow{p} 0$. Since $\hat{\boldsymbol{\theta}}$ minimizes $Q(\boldsymbol{\theta})$, $Q(\hat{\boldsymbol{\theta}}) \xrightarrow{p} 0$, which in turn implies

$$\mathrm{plim}_{n_m \to \infty}[\mathrm{tr}\{\mathbf{S}^{(t)}\boldsymbol{\Sigma}^{(t)-1}(\hat{\boldsymbol{\theta}})\} - \log|\mathbf{S}^{(t)}\boldsymbol{\Sigma}^{(t)-1}(\hat{\boldsymbol{\theta}})| - p^{(t)}] = 0$$

$$\mathrm{plim}_{n_m \to \infty}[\bar{\mathbf{z}}^{(t)} - \boldsymbol{\mu}^{(t)}(\hat{\boldsymbol{\theta}})]'\boldsymbol{\Sigma}^{(t)-1}(\hat{\boldsymbol{\theta}})[\bar{\mathbf{z}}^{(t)} - \boldsymbol{\mu}^{(t)}(\hat{\boldsymbol{\theta}})] = 0.$$

These two equations and Assumption 1 *i)* imply $\mathrm{plim}_{n_m \to \infty}\hat{\boldsymbol{\theta}} = \boldsymbol{\theta}_0$. To derive the limiting distribution of $\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$, we split $\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$ into two vectors

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) = \sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_{\mathbf{n}}) + \sqrt{n}(\boldsymbol{\theta}_{\mathbf{n}} - \boldsymbol{\theta}_0), \tag{4.9}$$

where

$$\boldsymbol{\theta}_{\mathbf{n}} = \begin{pmatrix} \boldsymbol{\tau}_0 \\ \bar{\mathbf{f}} \\ \mathbf{s}_{\mathbf{f}} \end{pmatrix}, \quad \bar{\mathbf{f}} = \begin{pmatrix} \bar{\mathbf{f}}^{(1)} \\ \vdots \\ \bar{\mathbf{f}}^{(T)} \end{pmatrix}, \quad \mathbf{s}_{\mathbf{f}} = \begin{pmatrix} \mathrm{vech}\mathbf{S}_{\mathbf{ff}}^{(1)} \\ \vdots \\ \mathrm{vech}\mathbf{S}_{\mathbf{ff}}^{(T)} \end{pmatrix},$$

$$\bar{\mathbf{f}}^{(t)} = \frac{1}{n^{(t)}}\sum_{i=1}^{n^{(t)}}\mathbf{f}_i^{(t)}, \quad \mathbf{S}_{\mathbf{ff}}^{(t)} = \frac{1}{n^{(t)}-1}\sum_{i=1}^{n^{(t)}}(\mathbf{f}_i^{(t)} - \bar{\mathbf{f}}^{(t)})(\mathbf{f}_i^{(t)} - \bar{\mathbf{f}}^{(t)})'.$$

Note that the $\boldsymbol{\tau}$-parts of $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}_{\mathbf{n}}$ are both $\boldsymbol{\tau}_0$, and that $\bar{\mathbf{f}}$ and $\mathbf{s}_{\mathbf{f}}$ in $\boldsymbol{\theta}_{\mathbf{n}}$ are sample moments of the unobservable factors. For the derivation of the limiting distribution of $\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_{\mathbf{n}})$, we expand the first derivative of $Q(\boldsymbol{\theta})$ evaluated at $\hat{\boldsymbol{\theta}}$ around $\boldsymbol{\theta}_{\mathbf{n}}$. The standard argument shows that

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) = (\mathbf{F}_0'\boldsymbol{\Omega}_0^{-1}\mathbf{F}_0)^{-1}\mathbf{F}_0'\boldsymbol{\Omega}_0^{-1}\sqrt{n}[\mathbf{c} - \boldsymbol{\gamma}(\boldsymbol{\theta}_{\mathbf{n}})] + o_p(1), \tag{4.10}$$

where

$$\mathbf{c} = \begin{pmatrix} \mathbf{c}^{(1)} \\ \vdots \\ \mathbf{c}^{(T)} \end{pmatrix}, \quad \mathbf{c}^{(t)} = \begin{pmatrix} \bar{\mathbf{z}}^{(t)} \\ \text{vech}(\mathbf{S}^{(t)}) \end{pmatrix},$$

$$\boldsymbol{\gamma}(\boldsymbol{\theta}) = \begin{pmatrix} \boldsymbol{\gamma}^{(1)}(\boldsymbol{\theta}) \\ \vdots \\ \boldsymbol{\gamma}^{(T)}(\boldsymbol{\theta}) \end{pmatrix}, \quad \boldsymbol{\gamma}^{(t)}(\boldsymbol{\theta}) = \begin{pmatrix} \boldsymbol{\mu}^{(t)}(\boldsymbol{\theta}) \\ \text{v}[\boldsymbol{\Sigma}^{(t)}(\boldsymbol{\theta})] \end{pmatrix},$$

and $\boldsymbol{\gamma}(\boldsymbol{\theta}_{\mathbf{n}})$ consists of, by (4.3),

$$\begin{aligned} \boldsymbol{\mu}^{(t)}(\boldsymbol{\theta}_{\mathbf{n}}) &= \boldsymbol{\beta}_0^{(t)}(\boldsymbol{\tau}_0) + \mathbf{B}^{(t)}(\boldsymbol{\tau}_0)\bar{\mathbf{f}}^{(t)}, \\ \boldsymbol{\Sigma}^{(t)}(\boldsymbol{\theta}_{\mathbf{n}}) &= \mathbf{B}^{(t)}(\boldsymbol{\tau}_0)\mathbf{S}_{\mathbf{ff}}^{(t)}\mathbf{B}^{(t)\prime}(\boldsymbol{\tau}_0) + \boldsymbol{\Psi}^{(t)}(\boldsymbol{\tau}_0). \end{aligned}$$

Note that, by (4.9),

$$\begin{aligned} \bar{\mathbf{z}}^{(t)} &= \boldsymbol{\beta}_0^{(t)}(\boldsymbol{\tau}_0) + \mathbf{B}^{(t)}(\boldsymbol{\tau}_0)\bar{\mathbf{f}}^{(t)} + \bar{\mathbf{e}}^{(t)}, \\ \mathbf{S}^{(t)} &= \mathbf{B}^{(t)}(\boldsymbol{\tau}_0)\mathbf{S}_{\mathbf{ff}}^{(t)}\mathbf{B}^{(t)\prime}(\boldsymbol{\tau}_0) + \mathbf{S}_{\mathbf{ee}}^{(t)} + \mathbf{B}^{(t)}(\boldsymbol{\tau}_0)\mathbf{S}_{\mathbf{fe}}^{(t)} + \mathbf{S}_{\mathbf{ef}}^{(t)}\mathbf{B}^{(t)\prime}(\boldsymbol{\tau}_0), \end{aligned}$$

where for any $\mathbf{u}$ and $\mathbf{v}$ $\mathbf{S}_{\mathbf{uv}}^{(t)}$ is defined as

$$\mathbf{S}_{\mathbf{uv}}^{(t)} = \frac{1}{n^{(t)}-1}\sum_{i=1}^{n^{(t)}}(\mathbf{u}_i^{(t)} - \bar{\mathbf{u}}^{(t)})(\mathbf{v}_i^{(t)} - \bar{\mathbf{v}}^{(t)})'.$$

Thus, the subvectors of $\mathbf{c} - \boldsymbol{\gamma}(\boldsymbol{\theta}_{\mathbf{n}})$ can be expressed as

$$\bar{\mathbf{z}}^{(t)} - \boldsymbol{\mu}^{(t)}(\boldsymbol{\theta}_{\mathbf{n}}) = \bar{\mathbf{e}}^{(t)}, \tag{4.11}$$

$$\begin{aligned} \text{vech}[\mathbf{S}^{(t)} - \boldsymbol{\Sigma}^{(t)}(\boldsymbol{\theta}_{\mathbf{n}})] &= \text{vech}[\mathbf{S}_{\mathbf{ee}}^{(t)} - \boldsymbol{\Psi}^{(t)}(\boldsymbol{\tau}_0)] + \mathbf{D}_{p^{(t)}}^{+}[(\mathbf{I}_{p^{(t)}} \otimes \mathbf{B}^{(t)}(\boldsymbol{\tau}_0)) \\ &\quad + (\mathbf{B}^{(t)}(\boldsymbol{\tau}_0) \otimes \mathbf{I}_{p^{(t)}})\mathbf{k}_{p^{(t)}p^{(t)}}]\text{vec}\mathbf{S}_{\mathbf{fe}}^{(t)}, \end{aligned} \tag{4.12}$$

where $\mathbf{k}_{mn}$, known as the commutation matrix, satisfies $\text{vec}\mathbf{A}' = \mathbf{k}_{mn}\text{vec}\mathbf{A}$ for an $m \times n$ $\mathbf{A}$. Since, for $m \times n$ $\mathbf{A}$ and $p \times q$ $\mathbf{B}$, $\mathbf{k}_{pm}(\mathbf{A} \otimes \mathbf{B}) = (\mathbf{B} \otimes \mathbf{A})\mathbf{k}_{qn}$, and since

$\mathbf{I}_{p^2} + \mathbf{k}_{pp} = 2\mathbf{D}_p\mathbf{D}_p^+$, it follows that

$$\text{vech}[\mathbf{S}^{(t)} - \boldsymbol{\Sigma}^{(t)}(\boldsymbol{\theta}_\mathbf{n})] = \text{vech}[\mathbf{S}_{\mathbf{ee}}^{(t)} - \boldsymbol{\Psi}^{(t)}(\boldsymbol{\tau}_0)] + 2\mathbf{D}_{p^{(t)}}^+[(\mathbf{I}_{p^{(t)}} \otimes \mathbf{B}^{(t)}(\boldsymbol{\tau}_0)]\text{vec}\mathbf{S}_{\mathbf{fe}}^{(t)}. \quad (4.13)$$

By (4.9), (4.10), (4.11), and (4.13), $\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_\mathbf{n})$ and $\sqrt{n}(\boldsymbol{\theta}_\mathbf{n} - \boldsymbol{\theta}_0)$ have a joint limiting normal distribution and are uncorrelated in the limit. Also, by (4.10), (4.11), and (4.13), the limiting distribution of $\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_\mathbf{n})$ is the same whether or not $\mathbf{f}_i^{(t)}$ are independent over t. Write this common limiting distribution as

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_\mathbf{n}) \xrightarrow{L} \mathrm{N}(\mathbf{0}, \mathbf{V}_1). \quad (4.14)$$

The limiting distribution of $\sqrt{n}(\boldsymbol{\theta}_\mathbf{n} - \boldsymbol{\theta}_0)$ is

$$\sqrt{n}(\boldsymbol{\theta}_\mathbf{n} - \boldsymbol{\theta}_0) \xrightarrow{L} \mathrm{N}(\mathbf{0}, \mathbf{V}_2),$$

where

$$\mathbf{V}_2 = \begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}_0 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{H}_0 \end{pmatrix}, \quad \mathbf{G}_0 = [\mathbf{G}_0^{(tm)}], \quad \mathbf{H}_0 = [\mathbf{H}_0^{(tm)}],$$

$$\mathbf{G}_0^{(tm)} = \lambda^{(tm)}\boldsymbol{\Phi}_0^{(tm)}, \quad t, m = 1, 2, \cdots, T,$$

$$\mathbf{H}_0^{(tm)} = \begin{cases} r^{(t)-1}2\mathbf{D}_{p^{(t)}}^+[\boldsymbol{\Phi}_0^{(tt)} \otimes \boldsymbol{\Phi}_0^{(tt)}]\mathbf{D}_{p^{(t)}}^{+\prime}, & \text{if } t = m, \\ \lambda^{(tm)}\mathbf{D}_{p^{(t)}}^+[\boldsymbol{\Phi}_0^{(tm)} \otimes \boldsymbol{\Phi}_0^{(tm)}]\mathbf{D}_{p^{(m)}}^{+\prime}, & \text{if } t \neq m. \end{cases}$$

The form of $\mathbf{V}_2$ is affected by the dependency of $\mathbf{f}_i^{(t)}$ over t, and we can write

$$\mathbf{V}_2 = \mathbf{V}_{2I} + \mathbf{V}_D,$$

where $\mathbf{V}_{2I}$ is $\mathbf{V}_2$ with all $\mathbf{\Phi}_0^{(tm)} = \mathbf{0}$ for $t \neq m$, and $\mathbf{V}_D$ is defined in the theorem. Hence, the limiting covariance matrix $\mathbf{V}$ of $\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$ is

$$\mathbf{V} = \mathbf{V}_1 + \mathbf{V}_2 = \mathbf{V}_I + \mathbf{V}_D,$$

where $\mathbf{V}_I = \mathbf{V}_1 + \mathbf{V}_{2I}$ is the limiting covariance matrix under the independence of $\mathbf{f}_i^{(t)}$ over t.

b) We use the results that

$$\mathrm{tr}[\mathbf{\Gamma}^{-1/2}\mathbf{C}\mathbf{\Gamma}^{-1/2} - \mathbf{I}]^2 = \mathrm{tr}[(\mathbf{C} - \mathbf{\Gamma})\mathbf{\Gamma}^{-1}]^2,$$

that $\mathrm{tr}(\mathbf{AB})^2 = \mathrm{vec}'\mathbf{B}(\mathbf{A} \otimes \mathbf{A})\mathrm{vec}\mathbf{B}$, and that if $\mathbf{A}_n$ is positive definite and $\mathbf{A}_n - \mathbf{I} = O_p(1/\sqrt{n})$ then

$$\log|\mathbf{A}_n| = \mathrm{tr}(\mathbf{A}_n - \mathbf{I}) - \frac{1}{2}\mathrm{tr}(\mathbf{A}_n - \mathbf{I})^2 + O_p(\frac{1}{\sqrt{n}}).$$

Then,

$$\begin{aligned}
Q(\hat{\boldsymbol{\theta}}) &= \sum_{t=1}^{T} n^{(t)}\{\mathrm{tr}[\mathbf{\Sigma}^{(t)-1/2}(\hat{\boldsymbol{\theta}})\mathbf{S}^{(t)}\mathbf{\Sigma}^{(t)-1/2}(\hat{\boldsymbol{\theta}}) - \mathbf{I}_{p^{(t)}}] - \log|\mathbf{\Sigma}^{(t)-1/2}(\hat{\boldsymbol{\theta}})\mathbf{S}^{(t)}\mathbf{\Sigma}^{(t)-1/2}(\hat{\boldsymbol{\theta}})| \\
&\quad +[\bar{\mathbf{z}}^{(t)} - \boldsymbol{\mu}^{(t)}(\hat{\boldsymbol{\theta}})]'\mathbf{\Sigma}^{(t)-1}(\hat{\boldsymbol{\theta}})[\bar{\mathbf{z}}^{(t)} - \boldsymbol{\mu}^{(t)}(\hat{\boldsymbol{\theta}})]\} \\
&= \sum_{t=1}^{T} n^{(t)}\{\frac{1}{2}\mathrm{tr}[[\mathbf{S}^{(t)} - \mathbf{\Sigma}^{(t)}(\hat{\boldsymbol{\theta}})]\mathbf{\Sigma}^{(t)-1}(\hat{\boldsymbol{\theta}})]^2 \\
&\quad +[\bar{\mathbf{z}}^{(t)} - \boldsymbol{\mu}^{(t)}(\hat{\boldsymbol{\theta}})]'\mathbf{\Sigma}^{(t)-1}(\hat{\boldsymbol{\theta}})[\bar{\mathbf{z}}^{(t)} - \boldsymbol{\mu}^{(t)}(\hat{\boldsymbol{\theta}})]\} + O_p(\frac{1}{\sqrt{n}}) \\
&= \sqrt{n}[\mathbf{c} - \boldsymbol{\gamma}(\hat{\boldsymbol{\theta}})]'\mathbf{\Omega}^{-1}(\hat{\boldsymbol{\theta}})\sqrt{n}[\mathbf{c} - \boldsymbol{\gamma}(\hat{\boldsymbol{\theta}})] + o_p(1).
\end{aligned}$$

Using $\boldsymbol{\theta}_{\mathbf{n}}$ in (4.9), we write

$$\sqrt{n}[\mathbf{c} - \boldsymbol{\gamma}(\hat{\boldsymbol{\theta}})] = \sqrt{n}[\mathbf{c} - \boldsymbol{\gamma}(\boldsymbol{\theta}_{\mathbf{n}})] + \sqrt{n}[\boldsymbol{\gamma}(\boldsymbol{\theta}_{\mathbf{n}}) - \boldsymbol{\gamma}(\hat{\boldsymbol{\theta}})].$$

By expanding $\boldsymbol{\gamma}(\hat{\boldsymbol{\theta}})$ at $\boldsymbol{\theta}_{\mathbf{n}}$, and by using (4.10), we obtain

$$\sqrt{n}[\boldsymbol{\gamma}(\boldsymbol{\theta}_{\mathbf{n}}) - \boldsymbol{\gamma}(\hat{\boldsymbol{\theta}})] = -\mathbf{F}_0(\mathbf{F}_0'\mathbf{\Omega}_0^{-1}\mathbf{F}_0)^{-1}\mathbf{F}_0'\mathbf{\Omega}_0^{-1}\sqrt{n}[\mathbf{c} - \boldsymbol{\gamma}(\boldsymbol{\theta}_{\mathbf{n}})] + o_p(1). \tag{4.15}$$

Thus,

$$Q(\hat{\boldsymbol{\theta}}) = \sqrt{n}[\mathbf{c} - \boldsymbol{\gamma}(\boldsymbol{\theta}_{\mathbf{n}})]'[\mathbf{I} - \boldsymbol{\Omega}_0^{-1}(\mathbf{F}_0'\boldsymbol{\Omega}_0^{-1}\mathbf{F}_0)^{-1}\mathbf{F}_0']\boldsymbol{\Omega}_0^{-1}\sqrt{n}[\mathbf{c} - \boldsymbol{\gamma}(\boldsymbol{\theta}_{\mathbf{n}})] + o_p(1).$$

It is known that the limiting distribution of $Q(\hat{\boldsymbol{\theta}})$ would be χ_q^2 distribution if the $\mathbf{f}_i^{(t)}$, $t = 1, 2, \cdots, T$ were independent. By (4.11) and (4.13), the limiting distribution of $\sqrt{n}[\mathbf{c} - \boldsymbol{\gamma}(\boldsymbol{\theta}_{\mathbf{n}})]$ is not affected by the dependency of $\mathbf{f}_i^{(t)}$ over t. Hence the result follows. □

Proof of Theorem 2

a) The parts of the proof of Theorem 1 not relying on the normality of $\mathbf{f}_i^{(t)}$ are valid for this theorem. Note that (4.10) and (4.14) hold without the normality of $\mathbf{f}_i$, since the limiting distribution of $\bar{\mathbf{e}}^{(t)}$, $\mathbf{S}_{\mathbf{ee}}^{(t)}$, and $\mathbf{S}_{\mathbf{fe}}^{(t)}$ in (4.11) and (4.13) does not depend on the normality of $\mathbf{f}_i$. Thus, the result follows by noting that $\mathbf{V}_I^{\boldsymbol{\tau}} = \mathbf{V}_1^{\boldsymbol{\tau}}$ the $\boldsymbol{\tau}$-part of $\mathbf{V}_1$ in (4.14).

b) Note that (4.15) holds also for this theorem. Thus, the argument used in the proof Theorem 1 b) shows the result. □

Proof of Theorem 3

a) Let $\boldsymbol{\theta}_{\mathbf{n}}^* = (\boldsymbol{\tau}_{10}', \mathbf{s}_{\mathbf{e}}', \bar{\mathbf{f}}', \mathbf{s}_{\mathbf{f}}')'$ where $\mathbf{s}_{\mathbf{e}} = (\mathbf{s}_{\mathbf{e}(1)}', \cdots, \mathbf{s}_{\mathbf{e}(T)}')'$, and $\mathbf{s}_{\mathbf{e}(t)}$ is the vector listing the diagonal elements of $\mathbf{S}_{\mathbf{ee}}^{(t)}$. By following the steps of the proof of Theorem 1 with $\boldsymbol{\theta}_{\mathbf{n}}^*$ in place of $\boldsymbol{\theta}_{\mathbf{n}}$, we can show that (4.10) holds with $\boldsymbol{\theta}_{\mathbf{n}}$ replaced by $\boldsymbol{\theta}_{\mathbf{n}}^*$. In this case, $\mathbf{c}^{(t)} - \boldsymbol{\gamma}^{(t)}(\boldsymbol{\theta}_{\mathbf{n}}^*)$ consists of (4.11) with $\boldsymbol{\theta}_{\mathbf{n}}$ replaced by $\boldsymbol{\theta}_{\mathbf{n}}^*$, and

$$\mathrm{vech}[\mathbf{S}^{(t)} - \boldsymbol{\Sigma}^{(t)}(\boldsymbol{\theta}_{\mathbf{n}}^*)] = \mathrm{vech}\mathbf{S}_{\mathbf{ee}}^{(t)*} + 2\mathbf{D}_{p(t)}^{+}[(\mathbf{I}_{p(t)} \otimes \mathbf{B}^{(t)}(\boldsymbol{\tau}_{10})]\mathrm{vec}\mathbf{S}_{\mathbf{fe}}^{(t)},$$

where the diagonal elements of $\mathbf{S}_{\mathbf{ee}}^{(t)*}$ are zero, and the off-diagonal elements are the same as those of $\mathbf{S}_{\mathbf{ee}}^{(t)}$. It follows that the limiting distributions of $\sqrt{n}[\mathbf{c} - \boldsymbol{\gamma}(\boldsymbol{\theta}_{\mathbf{n}}^*)]$, $\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_{\mathbf{n}}^*)$, and $\sqrt{n}(\hat{\boldsymbol{\tau}}_1 - \boldsymbol{\tau}_{10})$ do not depend on the normality of $\mathbf{f}_i$ and $\mathbf{e}_i^{(t)}$, nor on the independence of $\mathbf{f}_i^{(t)}$ over t. Thus, the general Assumption 2-b) does not alter the limiting distribution of $\sqrt{n}(\hat{\tau}_1 - \tau_{10})$ from the normal case.

b) It can be shown that (4.15) holds with $\boldsymbol{\theta}_{\mathrm{n}}$ replaced by $\boldsymbol{\theta}_{\mathrm{n}}^*$. Thus, the result follows from the fact, shown in the proof of part a), that $\sqrt{n}[\mathbf{c} - \gamma(\boldsymbol{\theta}_{\mathrm{n}}^*)]$ does not depend on the normality of $\mathbf{f}_i^{(t)}$ and $\mathbf{e}_i^{(t)}$, and on the independence of $\mathbf{f}_i^{(t)}$. □

5. CONCLUSION

The linear latent variable modeling can be useful for multi-population and longitudinal studies. Even for correlated populations and unbalanced data, and for cases involving fixed and non-normal latent variables, simple statistical procedures based on the assumption of normality and independence can be used for model fitting and inferences. The use of such methods has advantages. The analysis is simple and can be executed by the existing computer packages. The proper parameterization and model formulation provide meaningful interpretation and correct statistical inferences. Theoretical and numerical justification was given for the methods. The results expand the usefulness of the latent variable modeling in applied sciences, where non-normal, non-random, and unbalanced data in multi-population and longitudinal studies are often encountered.